Journal of Experimental Psychology

ARTHUR W. MELTON, Editor
Department of Psychology, University of Michigan
Ann Arbor, Michigan

DAVID A. GRANT, Associate Editor
University of Wisconsin

WILLIAM K. ESTES, Associate Editor
Indiana University

Consulting Editors

NORMAN H. ANDERSON, *University of California, Los Angeles*
E. JAMES ARCHER, *University of Wisconsin*
FRED ATTNEAVE, III, *University of Oregon*
CLETUS J. BURKE, *Indiana University*
JAMES DEESE, *Johns Hopkins University*
PAUL M. FITTS, *University of Michigan*
FRANK A. GELDARD, *University of Virginia*
JAMES J. GIBSON, *Cornell University*
CLARENCE H. GRAHAM, *Columbia University*
HAROLD W. HAKE, *University of Illinois*
ARTHUR L. IRION, *Tulane University*
HOWARD H. KENDLER, *New York University*
DONALD B. LINDSLEY, *University of California, Los Angeles*
KENNETH MACCORQUODALE, *University of Minnesota*
QUINN MCNEMAR, *Stanford University*
LEO POSTMAN, *University of California, Berkeley*
L. STARLING REID, *University of Virginia*
KENNETH W. SPENCE, *State University of Iowa*
DELOS D. WICKENS, *Ohio State University*

CONTENTS

Intermittent Reinforcement, Nonreversal Shifts, and Neutralizing in Concept Formation: I. GORMEZANO AND F. D. ABRAHAM 1

Three Types of Anchoring Effects in the Absolute Judgment of Hue: F. C. VOLKMANN AND T. ENGEN 7

Effects of Caffein on Enhancement in Foveal Simultaneous Contrast: J. P. KLEMAN, A. L. DIAMOND, AND E. SMITH 18

Recognition Memory for Nouns as a Function of Abstractness and Frequency: A. M. GORMAN 23

The Judgment of Size, Contrast, and Sharpness of Letter Forms: W. C. HOWELL AND C. L. KRAFT 30

The Subject Effect in Judgments of Subjective Magnitude: F. N. JONES AND M. J. MARCUS 40

Analysis of the Role of Overlearning in Discrimination Reversal: M. R. D'AMATO AND H. JAGODA 45

Figural Aftereffects in Kinesthesis: Effects of Object Width and Repeated Presentations: E. G. HEINEMANN 51

The Loci of Reinforcement: G. COLLIER AND L. MYERS 57

The Effects of Awareness on Verbal Conditioning: S. M. LEVIN 67

Proactive and Retroactive Effects of Overlearning: G. MANDLER AND C. K. KUHLMAN 76

Effects of Instructions on the Extinction of a Conditioned Finger-Withdrawal Response: R. H. LINDLEY AND K. E. MOYER 82

Information Processing Behavior: The Role of Irrelevant Stimulus Information: R. E. MORIN, B. FORRIN, AND W. ARCHER 89

American Psychological Association

Vol. 61, No. 1 January 1961

The JOURNAL OF EXPERIMENTAL PSYCHOLOGY is published monthly, two volumes per year, by the American Psychological Association, Inc., at Prince and Lemon Sts., Lancaster, Pa. The subscription rate per volume is $10.00, or $20.00 annually. Single copies are $2.00. Subscriptions, orders, address changes, and business communications should be addressed to the American Psychological Association, Inc., 1333 Sixteenth St. N.W., Washington 6, D. C.

This Journal publishes original experimental investigations which are intended to contribute toward the development of psychology as an experimental science. Studies with normal human subjects are favored over studies involving abnormal or animal subjects, except when the latter are specifically oriented toward the extension of behavior theory. Experimental psychometric studies and studies in applied experimental psychology or engineering psychology may be accepted if they have broad implications for experimental and theoretical psychology.

Normally, articles exceeding 20 printed pages cannot be accepted. Within this limit, the piecemeal experiment-by-experiment reporting of psychological research is discouraged, and the reporting of the data and conclusions of substantial segments of research efforts is encouraged. Specifically, an integrated series of studies accomplished simultaneously (e.g., most doctoral dissertations) *must* be presented in a single article. Special provision is made for the publication of Supplementary Reports and Replication Reports in articles of not more than 1.5 printed pages (see Editor's Note, this Journal, 1957, **53**, 1–2).

Address all articles submitted for publication to the Editor, Arthur W. Melton, Department of Psychology, University of Michigan, Ann Arbor, Michigan. **Manuscripts *must* adhere to the conventions concerning reference citations, preparation of tables and figures, manuscript format, etc., as described in the *Publication Manual of the American Psychological Association*.** When in doubt about practices of the Journal, authors should examine a recent issue.

Articles are published in the order of their receipt, except in rare circumstances. Authors are supplied with 50 free offprints without covers. All of the cost of an author's alterations in galley proof is charged to the author. Priority in publication is given to articles whose authors assume the full cost of publication. In 1961 the publication cost is $30 per printed page.

To save printing costs, supplementary material to articles in this Journal are deposited with the American Documentation Institute. Copies may be obtained by ordering the Document Number given in the footnote reference to ADI. Orders should be sent to ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C. Orders *must* be accompanied by checks or money orders payable to the Chief, Photoduplication Service, Library of Congress.

ARTHUR C. HOFFMAN
Managing Editor

HELEN ORR
Promotion Manager

Second-class postage paid at Lancaster, Pa.

Send address changes to 1333 Sixteenth St. N.W., Washington 6, D. C. Address changes must reach the Subscription Office by the 10th of the month to take effect the following month. Undelivered copies resulting from address changes will not be replaced; subscribers should notify the post office that they will guarantee second-class forwarding postage. Other claims for undelivered copies must be made within four months of publication.

© 1961 by the American Psychological Association, Inc.

Journal of
Experimental Psychology

VOL. 61, NO. 1 JANUARY 1961

INTERMITTENT REINFORCEMENT, NONREVERSAL SHIFTS, AND NEUTRALIZING IN CONCEPT FORMATION

ISIDORE GORMEZANO AND FRED D. ABRAHAM

Indiana University

The purpose of the present investigation was to determine the difficulty of sorting through the six stages of the Wisconsin Card Sorting Test (Grant & Berg, 1948), when for color and number sortings the irrelevant dimension of form was intermittently reinforced 0%, 25%, 50%, or 75% of the time.

Recent studies have attempted to assess the effects on concept formation of having intermittently reinforced irrelevant stimuli. Gormezano and Grant (1958) systematically varied the amount of intermittent reinforcement of an irrelevant dimension and determined its effect on the attainment of concepts on the WCST. Hammer (1955), as well as Gormezano and Grant (1958), investigated the transfer of irrelevant stimuli in a nonreversal shift situation (where the formerly irrelevant dimension becomes relevant). Kendler and D'Amato (1955), Buss (1956), and Harrow and Friedman (1958) concentrated on the possible role intermittent reinforcement of an irrelevant dimension might have on the observed superiority of reversal (where the relevant dimension stays the same but the positive and negative signs are reversed) over nonreversal shifts. None of these studies, however, concerned themselves with detailing the effects of having systematically varied the amount of intermittent reinforcement of the *same* irrelevant dimension through repeated shifts from one relevant dimension to another. The present investigation bears on this problem and is related to some recent formulations of Bourne and Restle (1959).

Bourne and Restle (1959) have extended Restle's discrimination theory (1955, 1957, 1958, 1959) to concept formation. According to the theory, the attainment of concepts involves two processes—the conditioning of relevant stimuli and the neutralizing of irrelevant ones. Irrelevant stimuli become neutralized with respect to more valid or relevant stimuli (those more highly correlated with reinforcement), and when the relevant stimuli are now made irrelevant, the neutralized irrelevant stimuli become deneutralized, regaining their disruptive effects. The

theory, however, fails to detail the role of degree of intermittent reinforcement of the irrelevant stimuli on the deneutralization process and the present study provides some parametric data on this point.

PROCEDURE

General procedure.—In the test situation a gray sorting tray with four double compartments was placed on a table before *S*. Four stimulus cards consisting of one red triangle, two green stars, three yellow crosses, and four blue circles were placed from left to right in the upper halves of the four double compartments. Six decks of response cards, differing in composition, were placed in a partitioned box fastened to one of the legs of the table, facing *E* and out of view of *S*. Each deck consisted of 48 response cards with each card containing one to four identical figures of a single color. There were four colors (red, green, yellow and blue) and four figures (triangles, stars, crosses and circles). Each card could therefore be sorted according to color, form, or number of figures.

The *S* was instructed in a standard manner (Gormezano & Grant, 1958). The *E* then selected a response card from one of the six decks and handed it to *S*. Initially, depending on the experimental condition, the category form, color, or number was called "correct," but after *S* had sorted 10 consecutive cards correctly according to the predetermined correct dimension, *E* then began selecting response cards from the second deck, shifting the correct category to a new dimension. After 10 correct sortings on this second dimension, *E* then began selecting response cards from a third deck. This procedure was repeated until *S* had satisfactorily sorted through each concept twice. (The *S*, therefore, had to shift his sorting behavior five times.) If *S* did not attain the criterion within 48 sortings in any of the six stages of the experiment, his data were discarded.

Experimental design.—The *S*s were assigned to one of four groups, each of which received six response-card decks. For a group, the composition of the six decks was such that when sorted on the basis of color or number, *form* was intermittently reinforced 0%, 25%, 50%, or 75% of the time, i.e., 0, 12, 24, or 36 of the 48 response cards in the "color" or "number" decks, when sorted on the basis of color or number and accordingly called "correct," would also have been "correct" on the basis of form. For the "form" decks, when *S* sorted on the basis of form, he received 0% reinforcement for color or number sortings.

There are 24 possible sequences of the three concepts in six stages if a concept does not follow itself, and if there are no repetitions until all three concepts have been attained. In the present experiment all 24 sequences were used to balance out practice and transfer effects. One *S* from each of the four experimental groups was randomly assigned to each of the 24 sequences, making 96 *S*s in all.

Subjects.—The *S*s were 96 men and women in elementary psychology courses at Indiana University. There were, in addition, 34 "failures" to complete all six stages of the experiment. Since these "failures" were distributed unsystematically over all experimental conditions they were discarded from the analysis. (There were 9, 6, 11, and 8 "failures" in the 0%, 25%, 50%, and 75% conditions, respectively.)

RESULTS

Each *S*'s performance at each stage of the experiment was recorded in terms of five scores: (*a*) the number of *perseverative errors* at each stage except the first, i.e., the number of incorrect responses which would have been correct on the immediately preceding stage; (*b*) the number of *nonperseverative errors*, i.e., the responses that would also have been wrong on the preceding stage of the experiment; (*c*) *total errors* at each stage of the experiment; (*d*) *total correct responses*, i.e., the number of correct sortings made by *S* at each stage of the experiment after he had made his first perseverative error and before he began the criterion sequence of 10 successive correct responses; (*e*) *total trials*, i.e., total number of sortings required to attain the criterion, excluding the 10 "criterion" cards.

The results of the experiment are summarized in Fig. 1 in which the averages of all scores, after a square-root transformation, are plotted against the six stages of sorting, with the percentage of reinforcement of the

irrelevant form dimension as the parameter. The top graph of Fig. 1 shows the average number of perseverative errors for the four experimental groups. By definition there can be no perseverative errors on the first stage. It will be noted that a considerable amount of perseveration takes place for the 25%, 50%, and 75% groups on the second stage. Relative to the control group these groups demonstrate a greater number of perseverative errors across the six stages, despite a marked drop in scores in the last three stages. The next lower graph shows the average number of nonperseverative errors at each stage of the experiment. On the first stage there is little difference among the four groups but on the third stage the 50% and 75% groups show a marked increase in the number of nonperseverative errors. In the fourth stage the nonperseverative errors of all groups decrease. This greater efficiency of performance is maintained in the subsequent stages of sorting.

The third part of Fig. 1 shows the average number of errors made at each of the six stages of sorting. Inspection of the graph fails to reveal any systematic trends among the four groups. The next lower graph in Fig. 1 shows the mean number of correct responses made at each of the six stages. The number of correct responses required to attain the criterion indicates the number of reinforcements or reminders needed to confirm or establish the correct sorting category and has been considered a measure of sorting difficulty (Grant & Curran, 1952). It can be seen that as intermittent reinforcement of the irrelevant form dimension was increased from 0% to 75%, the difficulty of sorting increases markedly. The difficulty of sorting is

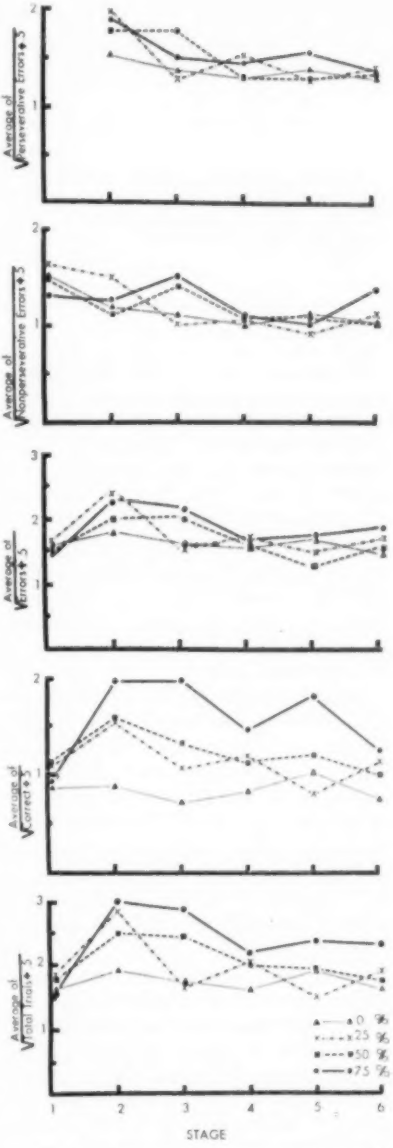

Fig. 1. Scores on successive stages with percentage of intermittent reinforcement of the irrelevant form dimension as the parameter.

TABLE 1

ANALYSES OF VARIANCE FOR DIFFERENT MEASURES OF PERFORMANCE

Source of Variation	*df*	*F* Ratios				
		Total Trials	Total Correct	Perseverative Errors	Nonperseverative Errors	Total Errors
1. Intermittent reinforcement of irrelevant form dimension	3	6.82**	18.60**	3.27*	.69	2.61
2. Initial-final sorting (Stages 1–3 vs. Stages 4–6)	1	4.76*	1.41	26.82**	28.11**	11.67**
3. Intermittent reinforcement and initial-final sorting	3	.59	.94	1.09	.42	.72
4. Error	568	(1.43)	(.73)	(.33)	(.36)	(.51)

Note.—Number in parentheses are error mean squares. The *df* for perseverative error scores is 472.
* $P = .05$.
** $P = .01$.

most pronounced for the 75% group and it will be noted that for this group there is no appreciable decrease in scores across the six stages. The bottom graph in Fig. 1 depicts the effect of 0%, 25%, 50%, and 75% reinforcement of the irrelevant form dimension on the mean number of trials. It can be seen that as intermittent reinforcement was increased from 0% to 75%, sorting becomes more difficult across all six stages of sorting.

Statistical analysis of each of the five scores presented in Fig. 1 was carried out by analysis of variance. Because all scores were frequency measures, it was considered desirable to use the square-root transformation $\sqrt{x + .5}$. This rational transformation serves to reduce the skewness of the frequency data, making them more amenable to the analysis of variance (Bartlett, 1947). Table 1 gives the summary of the analysis of variance for each of the scores. In Row 1, the source of variation tested is the effect of degree of intermittent reinforcement of the irrelevant dimension of form on the six stages of sorting. According to the *F* values there were significant differences in total trials, total correct responses, and perseverative errors. Row 2 shows that the shift between the first three sorting stages and the second three sorting stages resulted in a significant decrease in total trials, perseverative errors, nonperseverative errors and total errors. This represents the over-all learning that took place in the second half of the experiment vs. the first half. Only the *F* for total correct responses failed to reach significance. In Row 3 the interactions between Row 1 and Row 2 factors are tested. No significant interaction was found for any score.

DISCUSSION

There are two main findings in this investigation: (*a*) increasing the degree of intermittent reinforcement of the irrelevant form dimension during color and number sorting resulted in significant increases in total trials, total correct responses, and perseverative errors across the six stages; and (*b*) there was a significant decrease in the difficulty of sorting from the first three stages to the last three stages.

The first finding requires detailed consideration. The only basis *S* has for learning the correct sorting response is via the reinforcement or nonreinforcement of his responses. When reinforcement of the irrelevant dimension is increased it becomes increasingly difficult for *S* to learn whether the intermittently reinforced *irrelevant* dimension of form or the *relevant* dimension

of color or number is correct. Hence, increasing the percentage of reinforcement of the irrelevant dimension makes it increasingly difficult for *S* to learn which dimension is correct. This difficulty persists through repeated nonreversal shifts, even though the *same* irrelevant dimension is intermittently reinforced. Bourne and Restle (1959) are therefore quite correct in assuming that neutralized irrelevant stimuli regain their disruptive effects when the relevant stimuli with respect to which they have been neutralized, are now made irrelevant. Further, the present investigation indicates that the disruptive effects are an increasing function of the percentage of reinforcement of the irrelevant dimension.

The failure to find significant changes in nonperseverative and total errors while finding significant changes in perseverative errors requires detailed explanation. These results can be understood if it is noted that as percentage of reinforcement of form is increased there is less opportunity for *S* to make errors, i.e., at 0% reinforcement of the irrelevant form dimension each time *S* sorts on the basis of form an error is recorded whereas at 75% reinforcement if *S* "erroneously" sorts all cards on the basis of form, only 25% of the responses could be recorded as errors.

The present analysis of the data does not provide a clue for the significant effect of the independent variable on perseverative errors. An investigation recently reported by Buss (1959) suggests that in the present investigation it is a function of introducing form as the relevant dimension in two of the six stages of sorting. In the Buss study it was reported that increasing the percentage of reinforcement of a *previously* correct dimension results in a monotonic increase in difficulty of concept formation. In the present investigation this would best be reflected by an increase in perseverative errors. Unfortunately, our experimental design does not permit an orthogonal comparison of scores as a function of previous sorting on the relevant form dimension. However, a preliminary study by Abraham and Gormezano (1959) employing the WCST, reports findings substantially in agreement with Buss. A more extensive investigation attempting to detail this relationship is presently under way.

The second finding reflects an over-all learning that occurs in the last three stages of sorting. Once *S* has sorted through the first three stages of the WCST he has presumably learned which dimensions are relevant. This would be reflected in a decrease in the nonperseverative and total error scores. That this is not the only factor contributing to the over-all learning is indicated by the significant decrease in perseverative error scores. This measure suggests that a nonspecific type of learning may also be operating, i.e., learning how to learn (Harlow, 1949). In any case, the over-all learning indicates that the disruptive effects of the deneutralization process becomes attenuated in the later stages of sorting. Since the Restle and Bourne model assumes complete deneutralization this finding indicates the possibility that the model requires modification. Some aspects of Restle's learning set model (Restle, 1958) might well be employed for the task.

Summary

An investigation was made of the difficulty of sorting through the six stages of the Wisconsin Card Sorting Test when for color and number sortings the irrelevant dimension of form was intermittently reinforced 0%, 25%, 50%, or 75% of the time. A simple orthogonal design was employed. Twenty-four *S*s were assigned to each of the four conditions. The results and conclusions were as follows:

1. As intermittent reinforcement of the irrelevant form dimension was increased from 0% to 75%, sorting through the six stages became more difficult.

2. A decrease in sorting difficulty in the last three stages vs. the first three stages indicated that over-all learning had occurred.

3. It was concluded that in accord with Bourne and Restle's formulations (1959), neutralized irrelevant stimuli regain their disruptive effects when the relevant stimuli with respect to which they have been neu-

tralized, are now made irrelevant. Further, the disruptive effects are an increasing function of the percentage of reinforcement of the irrelevant dimension. However, these disruptive effects become attenuated by over-all learning and suggest the necessity of modifying the Bourne and Restle model.

REFERENCES

ABRAHAM, F. D., & GORMEZANO, I. Non-reversal shifts in concept formation with partial reinforcement. Paper read at Midwest. Psychol. Ass., Chicago, May 1959.

BARTLETT, M. S. The use of transformations. *Biometrics*, 1947, **3**, 39–52.

BOURNE, L. E., & RESTLE, F. Mathematical theory of concept identification. *Psychol. Rev.*, 1959, **66**, 278–296.

BUSS, A. H. Reversal and nonreversal shifts in concept formation with partial reinforcement eliminated. *J. exp. Psychol.*, 1956, **52**, 162–166.

BUSS, A. H. Supplementary report: Maintenance of a previously learned concept as a function of partial reinforcement. *J. exp. Psychol.*, 1959, **58**, 414–415.

GORMEZANO, I., & GRANT, D. A. Progressive ambiguity in the attainment of concepts on the Wisconsin Card Sorting Test. *J. exp. Psychol.*, 1958, **55**, 621–627.

GRANT, D. A., & BERG, E. A. A behavioral analysis of degree of reinforcement and ease of shifting to new responses in a Weigl-type card-sorting problem. *J. exp Psychol.*, 1948, **38**, 404–411.

GRANT, D. A., & CURRAN, J. F. Relative difficulty of number, form, and color concepts of a Weigl-type problem using unsystematic number cards. *J. exp. Psychol.*, 1952, **43**, 408–413.

HAMMER, M. The role of irrelevant stimuli in human discrimination learning. *J. exp. Psychol.*, 1955, **50**, 47–50.

HARLOW, H. F. The formation of learning sets. *Psychol. Rev.*, 1949, **56**, 51–65.

HARROW, M., & FRIEDMAN, G. B. Comparing reversal and nonreversal shifts in concept formation with partial reinforcement controlled. *J. exp. Psychol.*, 1958, **55**, 592–598.

KENDLER, H. H., & D'AMATO, M. F. A comparison of reversal shifts and nonreversal shifts in human concept formation behavior. *J. exp. Psychol.*, 1955, **49**, 165–174.

RESTLE, F. A theory of discrimination learning. *Psychol. Rev.*, 1955, **62**, 11–19.

RESTLE, F. Theory of selective learning with probable reinforcements. *Psychol. Rev.*, 1957, **64**, 182–191.

RESTLE, F. Toward a quantitative description of learning set data. *Psychol. Rev.*, 1958, **65**, 77–91.

RESTLE, F. Additivity of cues and easy to hard transfer in discrimination of consonant clusters. *J. exp. Psychol.*, 1959, **57**, 9–14.

(Received December 8, 1959)

Journal of Experimental Psychology
1961, Vol. 61, No. 1, 7-17

THREE TYPES OF ANCHORING EFFECTS IN THE ABSOLUTE JUDGMENT OF HUE[1]

FRANCES C. VOLKMANN AND TRYGG ENGEN

Brown University

In speaking of anchoring, psychologists have had several effects in mind. Each effect appears as a change in one or more dependent variables. Guilford has noted that ". . . the concept of an anchor . . . itself has needed anchoring and the psychological problem goes beyond the mere addition or use of a special stimulus called an *anchor*" (1954, p. 312). The purpose of this paper is to present a classification of anchoring effects, and to investigate experimentally three types of effects in the absolute judgment of hue.

The principal dependent variables which have been employed in experiments on judgment are frequencies of response, latencies of response, variability of response, and confidence attending responses. In a wide variety of experimental situations, usually in connection with a method of single stimuli, the measurement of these dependent variables has revealed one or more of the following effects, termed *anchoring effects*: (*a*) a set of correlated changes in response-frequency, known as a shift of a judgmental scale; (*b*) a selectively increased frequency of correct responses; (*c*) a selectively decreased latency of response; (*d*) a selectively decreased variability of responses; (*e*) a selectively increased confidence attending the responses of judging.

We can classify the various anchoring effects according to the operations by which they are produced. In this classification there are five major types of anchoring effects.

I. *Anchoring effects correlated with specific stimulus values.*—This class may be divided into two subclasses, *supplied-* and *end-*anchoring effects.[2] Anchoring stimuli may be *supplied* either within the range of stimuli to be judged or outside of this range. Reese, Reese, Volkmann, and Corbin (1953) studied the action of an anchoring stimulus placed within the range of judged stimuli. They found that if *S*s estimate the inclinations of lines representing radii of a circle, a visible inclination marker (a black line) placed into the continuum at 30° will greatly increase the accuracy of their judgments at and around that point. The action of an anchoring stimulus remote from the range of judged stimuli has customarily been considered in terms of a shift of a judgmental scale caused by the presentation of this stimulus. This shift has been noted by a number of workers (Heintz, 1950; Helson, 1948; Postman & Miller, 1945).

End-anchoring, the second type of stimulus anchoring, may be produced as a direct result of the range of the ordered classes of stimuli employed in a given experiment. For example, if *S* judges a finite range of ordered stimuli such as length of seen lines or weights of lifted objects, there is an anchoring effect on judgments of the longest and shortest line or the heaviest and lightest weight presented, and to a lesser degree on stimuli near them. End-anchoring effects have recently been studied in some detail by Eriksen and Hake (1955).

II. *Anchoring effects correlated with points on the continuum of a discriminable aspect.*—Two subclasses also fall under this heading, those of *natural-* and *derived-*anchoring effects. Natural anchoring is defined by the

[1] This study is based upon a thesis submitted by the first author to the Department of Psychology at Brown University in partial fulfillment of the requirements for the degree of Master of Arts. The authors wish to thank John Volkmann of Mount Holyoke College for his suggestions and criticisms concerning both the experiment and the manuscript.

[2] The classification of anchoring effects into the categories *supplied*, *natural*, and *derived* is that of Reese et al. (1953).

fact that under normal conditions certain anchoring effects appear which are independent of specific stimuli and the instructions of *E*. For example, *Ss*' judgments of inclination always show anchoring effects at the vertical and horizontal (Onley & Volkmann, 1958; Reese et al., 1953). These effects are to be differentiated from end-anchoring, where the anchoring point shifts as a function of the range of stimuli employed in a particular experiment.

Reese et al. (1953) have also noted that in judging inclination, *S*'s responses may show an anchoring effect at the 45° angle as well as at the horizontal and the vertical. The effect is not always noticeable, however. These authors called this a *derived* effect because of this fact and because of the reasonable inference that in anchoring at the 45° angle *S* is subjectively bisecting the angle between 0° and 90°.

III. *Anchoring effects correlated with attitude or instruction.*—Suppose that an *S* who is judging lifted weights is instructed to imagine some value of weight much heavier than any he is lifting. He is told to remember this value, and to let it define for him what he is to call "heavy." His judgments of actual stimulus weights are likely to be systematically shifted, as in the case of an actual stimulus anchor. This effect of *anchoring by a selected value* has been demonstrated for inclination (Volkmann, 1936) and affectivity (Hunt & Volkmann, 1937). Similar anchoring effects resulting from attitudes and value systems have been studied recently by Fensterhein and Tresselt (1953) and Sherif and Hovland (1953).

IV. *Anchoring effects correlated with specific responses.*—The role of response variables in absolute judgments has been studied by Hake and Eriksen (1956) and by Eriksen and Hake (1957). By using stimuli spaced around the circular hue continuum, these workers avoided end-anchoring effects produced by the stimuli. They then required their *Ss* to name each hue as it was presented with a given number from 1 to 20. Under these conditions *Ss* clearly selected numbers at the ends of the response continuum as anchoring agents, regardless of the stimuli to which they were attached. Thus with favorable conditions we may expect the responses on the ends of an ordered set of responses to produce anchoring effects, just as end stimuli may produce them. Moreover, it should be possible to supply anchoring responses to a set of existing responses just as anchoring stimuli are supplied to a set of judged stimuli. In short, we may conjecture that all of the effects correlated with specific stimulus values in Class I may also be correlated with the responses used by an *S* in judging.

V. *Anchoring effects correlated with a high degree of compatibility between a stimulus and a response.*—If *S* is asked to name each of a group of hue stimuli with a specific nonsense syllable, we would expect an anchoring effect if the syllable YEL were attached to the yellow hue, but not if it were attached to green or red, and not if a less compatible syllable were attached to the yellow hue (Fitts & Deininger, 1954).

The classification of anchoring effects is relatively broad; it opens new possibilities for the experimental investigation of anchoring. To take one possibility: we would like to find a single experimental situation in which several different types of anchoring effects could be produced at will. They could then be compared and even combined, in order to examine the more complex performances of judgment.

The present study attempts to manipulate both stimulus and response anchoring in a single experimental situation through the operations of: (*a*) presenting *S* with a class of stimuli and a class of responses which are as free of anchoring agents as *E* could make them; (*b*) inserting anchoring agents into these stimulus- and response-classes independently; and (*c*) comparing the resulting anchoring effects with one another and with the nonanchored condition.

Method

Stimuli and responses.—The basic stimulus class consisted of 20 Munsell hues equally spaced around the hue continuum. They were all of equal brightness and of equal and fairly low saturation. These hues, Numbers 5 and 10 5/4 of R, YR, Y, GY, G, BG, B, PB, P, and RP, provide a stimulus continuum which contains neither end-anchors nor obvious internal anchoring agents. Two more highly saturated hues, 10GY 5/8 and 10P 5/10, served as anchoring stimuli, replacing 10GY 5/4 and 10P 5/4, respectively, under conditions of stimulus anchoring.

The basic response class consisted of 20 nonsense syllables with association values of 20% or less (Hilgard, 1951): BIP, TIV, HEG,

DAX, PAF, GEX, YIF, MEZ, QUG, LEB, JAT, YIN, KEF, WUB, FEP, VAK, SEB, NUK, ZIC, and BOF. For conditions of response anchoring two syllables with high association value (DIC, 100% and JON, 93%) replaced two of the syllables of low association value.

Four different stimulus-response combinations permitted independent assessment of the following anchoring effects: (*a*) Associating the basic response class with the basic stimulus continuum should yield judgmental data which either show the effects of a nonanchored situation or point to natural anchoring agents. This condition will be denoted as stimulus-free, response-free (S-F, R-F), i.e., free of supplied anchoring agents. (*b*) Associating the basic response class with the stimulus continuum containing the two highly saturated hues should show the effect of stimulus anchoring agents. This condition will be termed stimulus-anchored, response-free (S-A, R-F). (*c*) Associating the response class containing the syllables DIC and JON with the basic stimulus continuum should show the effect of supplying anchoring agents to the responses. This combination will be designated as the S-F, R-A condition. (*d*) Associating the response class containing one anchoring syllable to the stimulus continuum containing one anchoring hue should show the relative effect of stimulus and response anchoring in the same situation. This condition will be denoted as S-A, R-A.

The following restrictions determined the manner in which these four stimulus-response combinations were used: (*a*) In order to separate the effects of stimulus and response anchoring, the experiment used 20 *Ss* in each of the four conditions; within a group every *S* received each syllable attached to a different hue. (The one exception to this procedure occurs in *d* below.) (*b*) Within a group each *S* received the syllables in a different order, a precaution against unknown order effects which might aid *Ss* in remembering particular syllables. (*c*) Within the limits of the above restrictions, each anchored response replaced each nonanchored response with as nearly equal frequency as possible. (*d*) In the S-A, R-A condition the same nonanchored stimulus always received the anchoring response. This permitted the plotting of group data with specific stimuli marked on the abscissa, showing the effect of stimulus-anchoring at one point and response-anchoring at another. (*e*) When the stimuli were presented to *S* for the first time in pre-experimental naming trials, the stimuli were given singly in order around the hue continuum. To balance for any possible end-anchoring effects in the group data which might result from this procedure, each *S* within a group received a different stimulus first. After the naming procedure the stimuli came in an irregular order. This restriction minimized end-anchoring effects in the judgments of individual *Ss*.

Apparatus.—The design of the apparatus enabled *E* to present each stimulus to *S* for a specified interval of time and to record *S*'s response and response time. The *S* sat directly in front of a gray viewing screen 40 in. long × 28 in. high which rested on a table. The stimuli appeared in a window in the middle of this screen at eye level. This viewing window and the screen around it were illuminated with 11.0 ft-L of light from a flood-lamp located behind and above *S*. A blue glass filter gave the light a daylight appearance.

The hue stimuli were presented as $\frac{3}{4} \times \frac{3}{4}$-in. chips mounted upon $4\frac{1}{2} \times 6$-in. cardboard backings which filled the viewing window and matched the gray color of the screen. Use of eight duplicate sets of stimuli minimized the possibility of identification by incidental marks or scratches which might not have been noticed by *E*.

When *E* dropped a stimulus card into a cardholder which was mounted behind the window in the screen, the bottom of the card triggered a microswitch attached behind the window, then came to rest on an aluminum base. The microswitch closed a telephone relay which had two independent output circuits. One circuit operated a synchronous timer which allowed the stimulus card to remain exposed for 3.5 sec., and then withdrew the aluminum base electronically to drop the card out of sight. The other circuit of the telephone relay closed a holding relay which started a stop clock. A simple voice-key system was designed to stop the clock and thus provide a measure of the time between the presentation of the stimulus and *S*'s response. The *S* spoke his responses into a microphone which fed an audioamplifier. The voice signal as amplified opened a sensitive rectifier-relay which then opened the holding relay and thus stopped the clock.

Subjects and procedure.—Eighty college sophomores, 48 men and 32 women, served as *Ss*. None of them were familiar with the purpose of the experiment, and all showed normal color vision on the Ishihara isochromatic test. Within each sex *Ss* were randomly assigned to the four experimental groups, with the restriction that each group contained 8 women and 12 men.

Upon arrival, *S* was first shown each hue

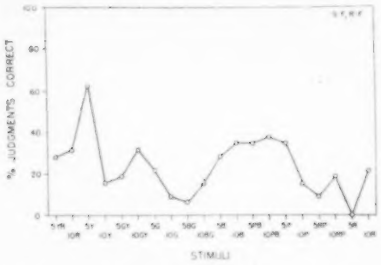

FIG. 1. Percentage of judgments correct as a function of stimuli for Group S-F, R-F, the control group.

around the continuum. The *E* pronounced and spelled the syllable which was to be associated with each stimulus, and *S* pronounced each syllable after *E*. During the remainder of the experimental session each hue was presented a total of 16 times in irregular order. As each stimulus appeared, *S* attempted to name it with the appropriate syllable as quickly and as accurately as possible. Although the stimulus could be viewed for 3.5 sec., *S* had approximately 5 sec. from the onset of the stimulus to make his response. At the end of this time *E* said "correct" if *S* had made the appropriate response. If *S* gave an incorrect response or no response *E* pronounced the correct one.

In employing the operations presented above, the present experiment differs in several ways from most previous experiments in the field of judgment: (*a*) The *S* is not instructed to discriminate a single discriminable aspect of the stimuli, but may use any discriminable aspect which the material offers him. (*b*) The experiment deals with nonordered response categories, rather than with ordered ones. (*c*) The *S* has to acquire the list of response categories during the course of the experiment. If he had been permitted to learn them previous to experimentation it would have been difficult to keep him from learning a chain of ordered classes. The procedure actually adopted simulates an experiment in the verbal learning of paired-associates. (*d*) The experiment permits comparison of several types of anchoring within the same experimental design.

RESULTS

Mean and median numbers of correct responses were calculated for each *S* in the four groups. Although both measures indicate the same effects in all cases, medians are reported since the distributions of individual scores are slightly positively skewed for low points and negatively skewed for high points. The Mann-Whitney *U* test evaluated differences between anchoring and nonanchoring conditions. Median response times were also computed and compared with the accuracy results.

Accuracy

Group S-F, R-F.—Figure 1 shows the percentage of judgments correct for each stimulus for Group S-F, R-F, the control group. Since each stimulus has been named with each syllable, the oscillations in the curve may be attributed to the stimuli only and not to the relative difficulty of the responses used. The curve as a whole contains two relatively high areas and two lower ones. The orange and yellow hues are more easily discriminated than the green, and the blues are more easily discriminated than are the purple and reddish hues. The hue 5Y is particularly noticeable. It was identified correctly about 25% more often than any other stimulus. A sign test shows judgments of the hue 5Y to differ significantly from

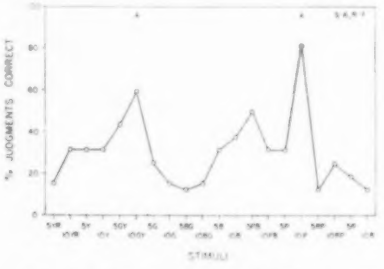

FIG. 2. Percentage of judgments correct as a function of stimuli for Group S-A, R-F. Arrows at top of graph indicate the loci of supplied anchoring stimuli.

those of half of the other points, i.e., $P < .01$, for 10 of 19 possible comparisons, and the Kendall coefficient of concordance, W, shows high agreement among *S*s in producing the curve as a whole ($P < .001$).

No other pronounced anchoring effects occur in the data of Group S-F, R-F. The individual data for the 20 *S*s comprising the group do not show end-anchoring effects at the stimuli and responses given them first and last in the preliminary naming procedure. Calculation of the percentage of judgments correct for each syllable showed them to vary from approximately 10% to 40%. This is similar to the range of correct identifications of the stimuli.

Group S-A, R-F.—Figure 2 presents the percentage of judgments correct for each stimulus for the group in which two highly saturated colors (anchoring stimuli) have replaced the hues of low saturation at 10GY and 10P. The curve as a whole resembles the corresponding curve for Group S-F, R-F, above. At the anchoring points, however, the accuracy of judgments shows a marked rise. The differences between each of these points and the corresponding points for Group S-F, R-F are significant ($P < .02$ and $P < .002$); the more highly saturated of the anchoring stimuli, 10P, produced the larger difference.[3] In addition, accuracy of identification of the hue 5Y has decreased, although this difference is not significant. These data thus provide clear evidence of supplied stimulus anchoring on the highly saturated hues.

Group S-F, R-A.—Figure 3 shows the results for the group for which two nonsense syllables of high meaningfulness replaced two of low meaningfulness. Since each label was attached to a different stimulus for each of the 20 *S*s, it was necessary to shift the abscissa one additional point for each successive *S* in order to place the anchoring responses at the same points in the group data. Thus, S_1 received an anchoring response at a hue corresponding to Point 10, S_2 received that response at Point 11, S_3 received it at Point 12, and so on for 20 *S*s. In Fig. 3, "anchor adjusted stimuli" means that each point on the abscissa represents every stimulus, and every point except Numbers 6 and 15 represents every nonanchored response. Points 6 and 15 represent every stimulus but only the anchored responses.

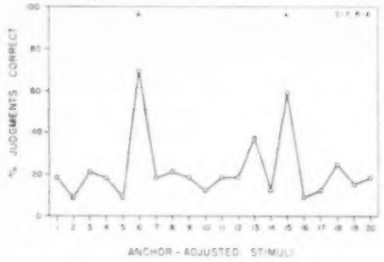

FIG. 3. Percentage of judgments correct as a function of anchor-adjusted stimuli for Group S-F, R-A. See text for further explanation of abscissa scale. Arrows at top of graph indicate the loci of supplied anchoring responses.

Since the abscissae are different, these data are not directly comparable to the data of Group S-F, R-F. Instead of comparing these points with the corresponding points in the control data, therefore, they are compared with the median percentage of correct judgments of the control group (25%). The differences between the points with response anchoring and the median of the control group are significant, with $P < .002$ in both cases, clearly demonstrating strong

[3] All P values reported are two-tailed.

response anchoring on the two syllables of high meaningfullness.

Group S-A, R-A.—This group was divided into two subgroups. The first subgroup, composed of the first 10 *Ss* who arrived, received an anchoring response at Hue 10GY and an anchoring stimulus at Hue 10P. The second 10 *Ss* received an anchoring stimulus at Hue 10GY and an anchoring response at Hue 10P. Figure 4 presents the results of both subgroups in terms of percentage of judgments correct at each stimulus. These curves are generally more variable than the curves for the three experimental groups previously discussed; this would be expected since the *N* is only half as large for each subgroup. Accuracy of identification is relatively high for both stimulus- and response-anchoring points, but in both subgroups the anchoring stimulus appears to have produced a much greater effect than the anchoring response. While the points with stimulus anchoring agents are highest on the curves of both subgroups, the response-anchored points were in both cases identified less accurately than at least one hue (5Y) which received no supplied anchoring agent. It should be noted that Hue 5Y is high in the S-F, R-F data also. In both cases the points with stimulus anchoring differ significantly from the corresponding points of the control group ($P < .002$ and $P < .055$ for hues 10P and 10GY, respectively). The differences with response anchoring are not significant in either case ($P > .10$), and no other significant differences occurred, although the curves as a whole are lower than for Group S-F, R-F.

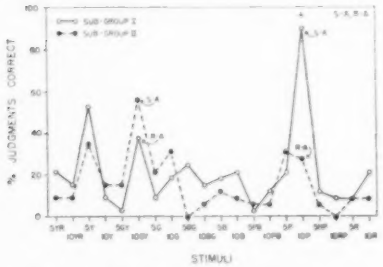

FIG. 4. Percentage of judgments correct as a function of stimuli for the two subgroups of Group S-A, R-A. The solid line and open circles represent the data of *Ss* who received an anchoring response at Hue 10GY and an anchoring stimulus at Hue 10P. The dotted line and solid circles show the data of *Ss* who received an anchoring stimulus at Hue 10GY and an anchoring response at Hue 10P.

The data as a whole suggest an interaction of anchoring effects when more than one type of anchor is presented in the same continuum. Response anchoring, though strong alone, is decreased when a supplied stimulus anchor is also given, at another point in the continuum. In addition, the high accuracy at Hue 5Y in the control group, which may suggest natural anchoring at this stimulus, appears to be decreased to a greater extent by the supplied stimulus anchor than by the supplied response anchor at 10GY in the subgroups of S-A, R-A. Accuracy of identification of Hue 5Y is still high in both subgroups, however. Thus, there is a general suggestion in the data that when these three types of anchoring are set into "competition," supplied stimulus anchoring is strongest, followed in order by "natural" stimulus anchoring[4] and finally by supplied response anchoring.

Response Times

Unfortunately, it was not possible to obtain a complete set of data on latency, because a number of the

[4] The term *natural* anchoring is used only tentatively here, and will be dealt with further in the discussion.

RTs were indeterminable. It will be recalled that the apparatus permitted measurement of latencies up to 5 sec. However, on some stimulus presentations some *S*s required longer than this to respond. This affects the RT data but not the accuracy data. Even with this limitation on the latency determinations, the data provided by the two measures look very similar. Differences between RTs in anchoring and nonanchoring conditions, evaluated by a less powerful median test, show the same trends as occur in the accuracy data. Both stimulus anchors in Group S-A, R-F produced significantly faster RTs than did the corresponding points in the control group ($P < .01$ at Hue 10GY and $P < .001$ at Hue 10P). In the response-anchored condition, median RTs at points with response anchoring were faster than for any other point, but the effects were not as pronounced as for stimulus anchoring, and are not significant. In the two subgroups of Group S-A, R-A the RT data again correspond closely to the accuracy data: differences are significant for both subgroups at points with stimulus anchoring ($P < .01$), and are high but not significantly so with response anchoring. Finally, the RT data support the accuracy results in indicating that Hue 5Y served as a "natural anchor" in Group S-F, R-F. The median RT at this stimulus was 2.02 sec. as compared with 3.22 for the other points, and a sign test showed 5Y to differ reliably ($P < .05$) from 10 of the remaining stimuli in Group S-F, R-F. No other significant differences occurred in any of the experimental groups.

In general, then, the data for RTs showed them to correspond to the accuracy data in showing both the existence of anchoring effects and the order of the interaction of these effects. In most cases, however, the differences are not as striking as those shown in the accuracy results.

Distributions of Responses for the Four Groups

Previous experiments (Eriksen & Hake, 1957; Reese et al., 1953) have indicated that an anchoring *stimulus* tends to increase the accuracy of judgments of stimuli on either side of it. There is the suggestion of this effect in the present data. Moreover, the data of the present experiment suggest that an anchoring *response* may produce slightly the opposite effect; it may act to decrease the accuracy of judgments of surrounding points on the curve. One might ask why these opposing effects are produced. A partial solution to this problem may be provided by calculating the frequency with which a response appropriate to a critical stimulus is attached to each stimulus in the continuum. Table 1 presents a representative sample of the results of these calculations. In this table, the anchoring points were chosen as the "critical stimuli" for each of the groups which received supplied anchoring agents. For the control group the same two points, Hues 10GY and 10P, were arbitrarily chosen to provide typical samples of the distributions in a relatively nonanchored situation. Actually this analysis was made for all points.

In Table 1, the critical stimulus (that stimulus for which a particular response was correct) is called "zero." The adjoining stimuli on either side of this stimulus are called "1" and "−1," and so on through the stimulus continuum. These numbers therefore represent the distance of each stimulus from the critical stimulus.

TABLE 1

FREQUENCY WITH WHICH A RESPONSE APPROPRIATE TO THE CRITICAL STIMULUS IS ATTACHED TO EACH STIMULUS IN THE CONTINUUM

Group	Critical Stimulus	Values of Stimuli																			N
		−9	−8	−7	−6	−5	−4	−3	−2	−1	0	1	2	3	4	5	6	7	8	9	10	
S-F, R-F																						
	10GY	0	0	1	0	0	1	0	9	47	98	31	30	18	4	1	0	0	0	0	0	240
	10P	0	1	0	2	0	1	0	0	8	69	43	12	10	0	2	0	0	0	1	0	149
S-A, R-F																						
	10GY	0	0	1	1	0	0	2	4	2	174	12	11	4	5	4	3	3	0	0	3	229
	10P	0	2	0	1	0	0	1	2	2	257	1	1	3	0	0	0	1	0	2	7	280
S-F, R-A																						
	JON	1	1	1	1	1	2	4	13	67	210	75	29	12	5	0	1	2	2	2	0	429
	DIC	2	1	1	1	1	3	8	21	83	192	64	23	5	1	0	0	0	0	0	0	406
S-A, R-A (1st half)																						
	10GY, R-A	0	0	0	0	0	1	3	8	34	77	52	33	7	2	3	1	2	0	0	1	224
	10P, S-A	0	1	0	0	0	1	0	0	0	125	0	1	0	1	0	0	0	0	0	0	128
S-A, R-A (2nd half)																						
	10GY, S-A	0	0	0	2	0	0	0	2	7	84	4	5	2	0	0	0	0	0	0	1	105
	10P, R-A	1	2	1	2	4	6	3	1	2	49	44	33	22	9	5	1	1	0	0	1	187

Note.—In the various groups the critical stimuli are as follows: S-F, R-F, the stimuli for the control group with the same hue and brightness as the anchoring stimuli; S-A, R-F, the anchoring stimulus; S-F, R-A, the stimulus appropriate to the anchoring response; S-A, R-A, one anchoring stimulus and one stimulus appropriate to an anchoring response. For further explanation see text.

The frequencies listed for each group indicate the total number of times that the response appropriate to stimulus "zero" was attached to each of the stimuli. Column N in the table shows the total number of times the response was used by each group of *S*s.

The response distributions for the two points sampled in the S-F, R-F condition show considerable spread to adjoining stimuli, and some spread as far as three stimuli in one direction. The distributions appear skewed in the clockwise direction. This may have been caused in part by originally presenting the stimuli in this order in the preliminary procedure. The distributions for Group S-A, R-F, however, are strikingly different. Although the anchoring stimulus was named correctly a much larger number of times, the frequency with which its response was attached to the stimuli on either side of it is almost negligible. This effect is also shown at the stimulus anchoring points in the two subgroups of condition S-A, R-A (Hue 10GY in the first half and Hue 10P in the second). Comparison of the distributions with stimulus anchoring to those with response anchoring reveals another striking difference. As shown for the S-F, R-A condition and for the response anchors of the S-A, R-A subgroups, response anchoring agents are attached with high frequency not only to the appropriate stimulus, but also to the adjoining stimuli. In general, the stimuli on either side of the correct one seem to be named by the anchoring response approximately a third as often as the response is used correctly. Some spread is shown at

stimuli two and three points removed from the critical one.

The effects of an anchoring stimulus and an anchoring response clearly appear to be different; an anchoring stimulus acts not only to produce a greater accuracy of response when it is presented, but acts also to separate the stimuli on either side of it from each other, and thus make them easier to discriminate. An anchoring response, on the other hand, does not alter the characteristics of the stimulus class at all. Consequently, if neighboring stimuli are relatively difficult to differentiate from each other *S*'s judgments show stimulus generalization rather than stimulus discrimination. He attaches the anchoring response, which is high in his response repertoire, not only to the appropriate stimulus, but to those surrounding it. A similar spread of responses without anchoring appears in the control data, and is consistent with the above interpretation.

The group data alone might suggest that the judgments of individual *S*s show anchoring effects only at supplied anchoring agents and possibly a natural stimulus anchor. However, an experimental situation free of all anchoring effects has probably not been found. Every *S* in the control condition showed indications of anchoring effects in his judgments of some few stimuli or responses. In most cases these anchoring effects were produced by different points for the different *S*s, and are therefore not apparent in the group data.

Discussion

Eriksen and Hake (1957), who also employed Munsell hues, mention in a footnote that the effects of their experimental variables were somewhat obscured by what appeared to be unequal discriminability among the stimuli. In the absence of an anchoring response near the greens, their *S*s showed least accuracy in naming Hues 5G through 10B and relatively high accuracy in the judgments of hues from 5PB to 5P, and at 10Y and 5GY. The present data reflect the same general trends. In addition, they suggest an anchoring effect at Hue 5Y in the control data. This hue was also judged with high accuracy in the Eriksen and Hake experiment but the accuracy was not as high as for some of the other non-anchored stimuli. Moreover, since both experiments used substantially the same group of stimuli, invariant anchoring at this point has not been tested. At the present time, therefore, we may conclude only that the data clearly suggest the possibility of a natural anchoring agent at Hue 5Y. There is at present no obvious way of determining whether there are natural anchoring agents in other regions of the hue continuum.

The finding of the present experiment, that supplied anchoring stimuli clearly produce consistent anchoring effects in the judgment of hue, supports similar findings by Volkmann (1936) and by Reese et al. (1953), in judgments of inclination. Moreover, the present experiment demonstrates these effects in the absence of any special anchoring instructions. An increase in saturation alone is sufficient to produce supplied anchoring effects.

Comparable anchoring effects produced by supplied response anchors have not been demonstrated previously. The occurrence of these effects in the present experiment, however, is consistent with the occurrence of end anchoring produced by using an ordinal response class (Eriksen & Hake, 1957). The results of the present experiment demonstrate not only that supplied response anchors are effective, but also that their effect occurs in the absence of order in the response class.

A comparison of stimulus and response anchoring shows them to differ markedly both in their relative strength and in the

manner in which they affect judgments of neighboring stimuli. One might ask, then, how these various and strikingly different anchoring effects may be explained by a theory of anchoring. Two major theoretical attempts to account for anchoring effects have been proposed: the adaptation-level hypothesis of Helson (1948), and the subjective-standard hypothesis of Eriksen and Hake (1957). The former hypothesis was constructed (or extended) to account for shifts in a judgmental scale which result from remote anchoring stimuli. The latter hypothesis was designed to explain end anchoring produced by either stimuli or responses. Anchoring effects, consequently, are explained in terms of a weighted averaging effect that occurs among a group of stimuli, or in terms of the action of the anchoring agent as a standard, in relation to which other judgments are made. Isolated, nonspreading anchoring effects shown in the present results may pose a problem for these theories.[5] The theories, particularly the latter one, are too vaguely constructed to allow us to determine precisely how they might account for these rather different effects. The present results may also suggest a quite different direction in which anchoring theory might usefully proceed.

Judgments are a class of responses evoked in humans by discriminative stimuli. The basic problems of anchoring, when considered in this manner, are the problems of stimulus discrimination and generalization, of response acquisition, and of stimulus-response association. Thus, the basic problems of anchoring and of judgment are also the basic problems of behavior theory. In this experiment in particular, the distributions presented in Table 1 for stimulus anchoring remind one forcibly of the results of instituting a discrimination in animals by the operations of pairing an S^D with an S^Δ. Both curves show a peak of responses to a narrow band of stimuli with very few responses to neighboring ones. On the other hand, response anchoring curves resemble initial generalization curves, which occur previous to the sharpening of discrimination by differential reinforcement.

Thus, anchoring effects need to be related more specifically to behavior theory in general. In terms of the older rubrics of psychology, the present experiment may be regarded either as an experiment in judgment or an experiment in learning, for the two fields share the same basic problems. It is unlikely that they will be solved in either of these fields alone.

[5] Parducci (1959) has dealt with some problems posed for adaptation-level theory by stimulus discriminability in a scaling task.

Summary

A framework of a definition and classification of judgment anchoring is presented and experimental evidence concerning two major types of anchoring, supplied stimulus and response anchors, is reported. Through the use of desaturated Munsell colors as stimuli and nonsense syllables of low meaningfulness as responses it was possible to obtain a more nearly anchorless experimental situation than has been used in previous experiments.

Two supplied anchoring stimuli with relatively high saturation values were identified with greater accuracy than the other 18 desaturated stimulus hues used in the experiment. There was only a slight indication that such anchor stimuli produced an increase in correct identifications of stimuli lying on either side of the anchors on the circular Munsell scale, suggesting the possibility that the use of a nonordered response class may have served to minimize the spread of anchoring effects.

Anchoring effects may also be produced by supplied anchoring responses. Two nonsense syllables of relatively high association values produced a significant increase in accuracy of identification of the stimuli to which these response anchors were attached. The data suggest that response anchoring may decrease the accuracy of identification of surrounding stimuli. If the stimuli are difficult to discriminate, *S* appears to use the anchoring response not only for the correct stimulus, but also for stimuli adjoining it.

Thus one difference between stimulus anchoring and response anchoring may lie in the differential effects which they produce at adjoining points. Moreover, when an anchoring stimulus and an anchoring response were presented at different points but in the

same situation, the effect of the anchoring stimulus was stronger.

Reaction speeds increased at anchoring stimuli, and similar, but less pronounced effects were suggested for response anchors.

The discussion of the results suggests that anchoring effects are related to behavior theory.

References

ERIKSEN, C. W., & HAKE, H. W. Absolute judgments as a function of stimulus range and number of stimulus and response categories. *J. exp. Psychol.*, 1955, **49,** 323–332.

ERIKSEN, C. W., & HAKE, H. W. Anchor effects in absolute judgments. *J. exp. Psychol.*, 1957, **53,** 132–138.

FENSTERHEIN, H., & TRESSELT, M. E. The influence of value systems on perception of people. *J. abnorm. soc. Psychol.*, 1953, **48,** 93–98.

FITTS, P. M., & DEININGER, R. S-R compatibility: Correspondence among paired elements within stimulus and response codes. *J. exp. Psychol.*, 1954, **48,** 483–492.

GUILFORD, J. P. *Psychometric methods.* New York: McGraw-Hill, 1954.

HAKE, H. W., & ERIKSEN, C. W. Role of response variables in recognition and identification of complex visual forms. *J. exp. Psychol.*, 1956, **52,** 235–243.

HEINTZ, R. K. The effect of remote anchoring points upon the judgment of lifted weights. *J. exp. Psychol.*, 1950, **40,** 584–591.

HELSON, H. Adaptation-level as frame of reference for prediction of psychophysical data. *Amer. J. Psychol.*, 1947, **60,** 1–29.

HILGARD, E. R. Methods and procedures in the study of learning. In S. S. Stevens (Ed.), *Handbook of experimental psychology.* New York: Wiley, 1951. Ch. 15.

HUNT, W. A., & VOLKMANN, J. The anchoring of an affective scale. *Amer. J. Psychol.*, 1937, **49,** 88–92.

ONLEY, J. W., & VOLKMANN, J. The visual perception of perpendicularity. *Amer. J. Psychol.*, 1958, **71,** 504–516.

PARDUCCI, A. An adaptation-level analysis of ordinal effects in judgment. *J. exp. Psychol.*, 1959, **58,** 239–246.

POSTMAN, L., & MILLER, G. A. Anchoring of temporal judgments. *Amer. J. Psychol.*, 1945, **58,** 510–518.

REESE, E. P., REESE, T. W., VOLKMANN, J., & CORBIN, H. H. (Eds.) Psychophysical research summary report, 1946–1952. *USN Special Devices Center tech. Rep.*, 1953, No. 131-1-5 (NAVEXOS P-1104).

SHERIF, M., & HOVLAND, C. I. Judgmental phenomena and scales of attitude measurement: Placement of items with individual choice of number of categories. *J. abnorm. soc. Psychol.*, 1953, **48,** 135–141.

VOLKMANN, J. Anchoring of absolute scales. *Psychol. Bull.*, 1936, **33,** 742–743.

(Received November 16, 1959)

Journal of Experimental Psychology
1961, Vol. 61, No. 1, 18–22

EFFECTS OF CAFFEINE ON ENHANCEMENT IN FOVEAL SIMULTANEOUS CONTRAST[1]

J. P. KLEMAN, A. L. DIAMOND, AND ESTHER SMITH

Psychological Research Center, University of Hawaii

Simultaneous brightness contrast is a visual phenomenon which takes place when the brightness of a test field is affected by the stimulation of an adjacent or surrounding inducing field. Test brightness can be affected in two ways: enhanced or depressed. Enhancement has been shown to occur when the inducing luminance is less than test luminance (Heinemann, 1955). Depression occurs when the inducing luminance is equal to or greater than the test luminance (Diamond, 1953; Leibowitz, Mote, & Thurlow, 1953).

The purpose of the present experiment is to determine what effect the administration of caffeine would have upon the contrast function. We speculated that caffeine could affect such a visual function since it is known that caffeine does stimulate certain medullary centers (Davison, 1949, p. 302) and, secondly, that when the reticular formation, in the medulla, is electrically excited, retinal activity increases (Granit, 1955, p. 107). Such information of course does not afford an exact hypothesis with respect to the effect of caffeine upon the contrast function but does make a caffeine investigation somewhat promising.

This experiment studied the simultaneous brightness contrast function under two conditions, caffeine administration and placebo administration, at three different levels of test-field luminance and a wide range of inducing luminances (below and above test luminances).

[1] This investigation was supported in part by a PHS research grant, B-1530, from the National Institute of Neurological Diseases and Blindness, United States Public Health Service.

METHOD

Apparatus.—The apparatus is shown schematically in Fig. 1. The basic features are the following: (*a*) binocular matching apparatus and necessary baffling are shown in Part A of Fig. 1; (*b*) the 2 × 2-in. glass slides with neutral-density gelatin Wratten filter with match field and test-inducing field patterns. The construction of these slides and graph of these patterns are presented in Part B of Fig. 1; (*c*) a series of 2 × 2-in. glass Wratten filters; and (*d*) optician's trial frames with mounted circular prisms.

The light source for each path is a 100-w., 120-v., tungsten filament bulb. In front of each bulb is a section of diffusing glass mounted over the proximal portion of the filter holders. The luminance at the slide holder is 350 ft-c as determined by a Weston foot-candle meter. This foot-candle measure converts to a metric measure of 367.6 mL. which provide a maximum luminance at the eye of 2.58 log mL.

A chin rest is provided to keep *S*'s head steady and at a constant distance from the matching apparatus. The distance between the chin rest and the two visual patterns is 122 cm. and the diameter of the test-inducing patterns, as seen by the right eye, is 1.1 cm. which provides an exact visual angle of 30′. The right fixation point is located 15′ to the left of the nearest edge of the inducing field. The pattern seen by the left eye is 4 mm. in diameter and a fixation point is located 15′ to the right of the nearest edge of this pattern. In both instances the stimulation may be considered essentially foveal since the visual angle between a fixation point and the furthest edge of its adjacent pattern does not exceed 45′.

The center baffle is provided so that the right eye sees only the right pattern and the left eye sees only the left pattern. An additional purpose of the box baffling attached to

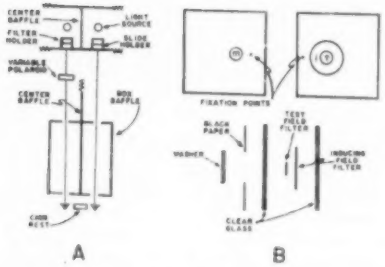

FIG. 1. (A) The schema of the apparatus; (B, top) front view of match-field slide and inducing-test-field slide; (B, bottom) schema of slide construction.

the center is to protect *S*'s eyes from stray light which might affect the brightness response being measured.

The inducing field densities range from 2.0 logarithmic units more dense than the test field to 0.1 logarithmic units less dense than the test field. The filters making up the test-inducing field patterns are sealed between two pieces of 2 × 2-in. clear glass. A piece of black paper with a hole 13 mm. in diameter is placed on the outside of the clear glass over the pattern. A metal washer is placed on top of this to give the pattern a distinct circle and to give a sharp edge. When additional glass filters are added behind the slides they affect both fields equally.

The construction of the slides for the match field is identical with the construction of the test-inducing pattern slides except no inducing field is included and therefore no metal washer is necessary. It is possible to vary the filter densities of the match field to cover a luminance range from no light to +2.58 log mL. In the left visual path there is also a variable density polaroid which can be operated by *E* such that the luminance of the match field can be varied continuously. The calibration of the variable polaroid within the range of 0.00 to 0.70 logarithmic units showed the cosine squared law of crossed polaroids to obtain within this range.

Neutral Wratten glass filters are used in both visual paths in order to control luminance values in large discrete steps. In the right visual path they change test- and inducing-field luminances simultaneously. In the left visual path they are used in conjunction with the variable polaroid, to adjust the luminance of the match field.

The circular prisms mounted in the optician's trial frames are manipulated by *S* to develop the binocular view shown in Fig. 2B.

Procedure.—The view presented to *S* is shown in Fig. 2. The apparatus is designed in such a way that the match field (m) can be set equal to the apparent brightness of the test field (t). The inducing field (i) can vary around the test field. The filter densities in the left visual path plus the density readings from the calibrated variable polaroid provide a measure of the apparent brightness of the test field. In this experiment measures were obtained using a form of the psychophysical method of limits; the method of equivalents (Guilford, 1954, p. 111). It has been demonstrated (Diamond, Scheible, Schwartz, & Young, 1955) that the results obtained in simultaneous contrast experimentation do not differ significantly when *E* uses either the psychophysical method of limits, constant stimuli, or adjustment.

On each of two days *S* went through both conditions. The first was without caffeine (Cond. NC) and the second was with caffeine (Cond. C). In each condition *S* was asked to make 10 matches at each of four or six points of inducing-field luminance which were varied above and below a constant test-field luminance, for three different levels of test-field luminance (see Table 1 for values). With this system it was possible to experiment at luminance levels where enhancement has previously been found (e.g., Heinemann, 1955). The possibility of counterbalancing Cond. C and NC was considered. Since the effect of caffeine upon enhancement (found in this experiment) lasted up to 6 hr., the usual methods of counterbalancing were not employed. Instead we conducted an exploratory study with each *S*, which demonstrated that when caffeine was not administered in both of two sessions, one following the other, an equal amount of enhancement occurred in the first and second sessions.

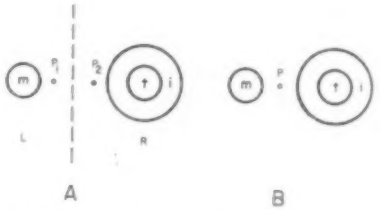

FIG. 2. (A) Patterns L and R as seen monocularly by *S*—L is presented to the left eye only, R to the right eye only; (B) the binocular view when Fixation Points P_1 and P_2 are fused.

This means that any change in enhancement that might occur as the result now of caffeine administration (Cond. C) is not very likely to be an order effect from a previous condition of no caffeine (Cond. NC).

The three *Ss*, male college students, were told that the experiment was attempting to measure the effects of a stimulant, caffeine, on a particular visual phenomenon, and they would be asked to take approximately the same amount of caffeine as that contained in two cups of coffee. For this experiment *Ss* were chosen who used caffeine-containing beverages only moderately or not at all. One *S* abstained altogether. All *Ss* were instructed not to use these beverages within 24 hr. previous to the experiment. Since in an exploratory experiment any visual effect attributed to caffeine administration lasted no longer than 6 hr., 24 hr. seemed reasonable.

Each experimental session (one sitting) lasted approximately 1½ hr.; about 45 min. were required for each of the two conditions. In Cond. NC, *S* was given a placebo identical with the capsule containing the caffeine. This was followed by 15 min. of dark adaptation in a light proof room. In the same room *S* was then asked to manipulate the circular prisms, mounted in the optician's trial frames, so as to fuse Fixation Points P_1 and P_2 and develop the pattern shown in Fig. 2B. It was possible to provide a period of light adaptation of at least 1.5 min. by having *S* make four "practice" matches. This was done at every point. The "practice" matches were followed by 10 more matches which provided the experimental data. When *E* was changing the inducing-field luminances by changing slides and/or filters *S* was dark-adapted for at least 2 min. This procedure was continued until a complete set of data was collected for each point of inducing-field luminances. Inducing luminances were presented in randomized fashion.

After the completion of Cond. NC, *S* was given 3 grains of anhydrous caffeine. This was followed by a 15-min. rest period and in turn followed by another 15 min. of dark adaptation. The procedure in Cond. C was identical with the procedure in Cond. NC.

In a previous exploratory study in which tension and/or anxiety were aroused by electrical shock, the amount of enhancement diminished. The results of this preliminary study made it advisable to eliminate testing of *Ss* during periods of examinations or at other times when they expressed feelings of discomfort or tenseness.

TABLE 1

EFFECT OF CAFFEINE ON APPARENT BRIGHTNESS OF TEST FIELD (B_m) IN LOG MILLILAMBERTS AS A FUNCTION OF INDUCING-FIELD LUMINANCE (B_i), FOR THREE DIFFERENT TEST-FIELD LUMINANCES (B_t)

Log B_i	Log B_m: No Caffeine			Log B_m: Caffeine		
	*S*JB	*S*TH	*S*EB	*S*JB	*S*TH	*S*EB
	Test-Field Luminance (Log B_t) = +1.08					
$-\infty$	1.05	1.04	0.99	1.06	1.03	1.00
−0.92	1.10	1.16	1.12	1.00	0.94	0.95
0.08	1.13	0.90	1.06	0.63	0.78	0.73
0.48	1.16	0.79	0.90	0.49	0.55	0.62
0.78	0.35	0.56	0.64	0.19	0.30	0.49
1.18	−1.10	−1.73	−1.50	−1.12	−1.72	−1.51
	Test-Field Luminance (Log B_t) = −0.42					
$-\infty$	−0.38	−0.43	−0.56	−0.40	−0.45	−0.53
−1.02	−0.22	−0.31	−0.43	−0.55	−0.56	−0.61
−0.82	−0.33	−0.52	−0.51	−0.62	−0.64	−0.71
−0.32	−1.35	−1.60	−2.32	−1.34	−1.54	−2.27
	Test-Field Luminance (Log B_t) = −1.42					
$-\infty$	−1.37	−1.23	−1.43	−1.32	−1.21	−1.45
−2.02	−1.20	−1.11	−1.34	−1.64	−1.44	−1.62
−1.82	−1.27	−1.15	−1.51	−1.72	−1.57	−1.79
−1.32	−1.86	−1.74	−2.12	−1.91	−1.73	−2.06

Note.—All figures represent means of data gathered from two experimental sessions at each of three test-field luminances.

RESULTS

The data are given in Table 1. In Fig. 3 the brightness of the test field (log B_m) is plotted as a function of the log luminance (log B_i) of the inducing field. Curves are plotted for Cond. C and NC for each of the three test-field luminances for each of the three *Ss*. Each point represents the mean of 20 settings (10 on each of two days) of the match-field luminance. These means represent the brightness of the test field.

It can be seen that all Cond. NC measures show the enhancement phenomena reported by Heinemann (1955). In each instance the brightness of the test field, recorded by *S*, first increased as the inducing-field luminance increased from zero. The brightness of the test field then decreased as the luminance of the

inducing field approached the luminance of the test field. At the point where the luminance of the inducing field is 0.1 logarithmic units greater than the test field or at its brightest, the brightness of the test field is decreased to its lowest point.

In Cond. C the effects of the caffeine not only diminish the enhancement phenomenon but erase it altogether so that the brightness of the test field is depressed in direct proportion to the luminance of the inducing field. Furthermore, the shapes of the curves for the data of each *S* have similar appearances. If we disregard the first and last points in each curve of Fig. 3 (for inducing luminance of 0 mL. and some maximum amount) and pay attention only to those points which show the effect of caffeine upon enhancement, i.e., those inducing luminance values at which the caffeine (Cond. C) curve drops below the control (Cond. NC), we find little or no overlap between the readings of the two curves. Actually, for these points, the mean values of Cond. C are significantly different from Cond. NC beyond the .001 level.

It can be seen that the two extremes of the function do not change from drug to nondrug state. One extreme is where the luminance of the inducing field is 0.1 logarithmic units greater than the test field. The other extreme is at the beginning of the curve where the inducing field is of infinite density. At neither point does the brightness of the test field differ significantly from control to experimental situation.

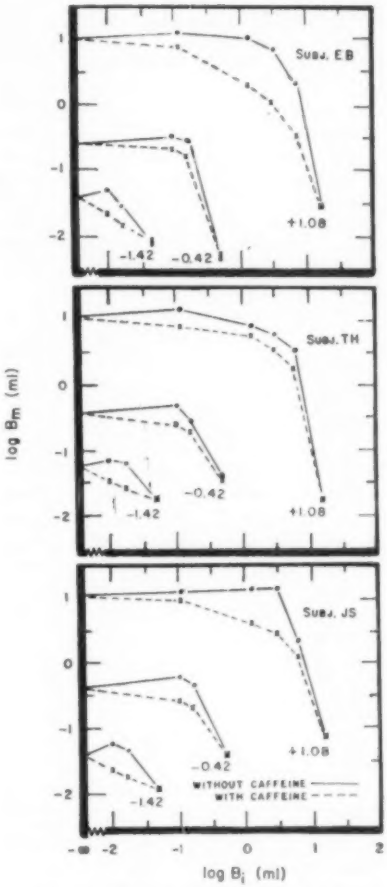

FIG. 3. Test-field apparent brightness (B_m) as a function of inducing-field luminance (B_i) with and without caffeine. The parameters on the curves are the test-field luminances in log millilamberts.

DISCUSSION

This experiment attempted to measure the effect of caffeine upon psychophysical enhancement. It can be seen in Table 1 and Fig. 3 that enhancement does not occur with caffeine administration. In order to understand why caffeine affects enhancement we should first explore the possible mechanism for enhancement itself. Diamond (1960) has suggested that enhancement occurs, in a stimulus situation such as ours, because the "on" retinal fibers stimulated within the test field are released from the inhibitory effects of "off" fiber discharge in the

inducing field, when the "off" fibers are themselves depressed by the inducing-field light. Granit describes such "on" and "off" fibers and further shows that an antagonism exists between them (Granit, 1955, pp. 69–72). According to this theory, enhancement comes from a disinhibition of the test-field fibers when the "off" fibers in the inducing field are depressed. Disinhibition, incidentally, has been demonstrated in the *Limulus* retina (Hartline & Ratliff, 1957).

With respect to the mechanism responsible for the effects of caffeine upon enhancement, we can only speculate as mentioned above. It is known that caffeine has a stimulant effect on the medulla (Davison, 1949, p. 302) and that stimulation of the reticular formation in the medulla increases both "on" and "off" fiber activity in the retina (Granit, 1955, p. 107). If the above theory of enhancement is correct, then it is possible that caffeine is somehow affecting enhancement through stimulation of the reticular formation and subsequent excitation of the "on" and "off" fibers. Whether or not this is the case of course will depend upon further experimentation.

Summary

In an experimental contrast situation, in which a small, circular test field is surrounded by an inducing annulus, the luminance of the inducing field was varied from zero to approximately 15 mL. This was done for three different levels of test-field luminances and for two conditions of drug administration. In one condition *Ss* were given 3 grains of caffeine. In the second condition they were given a placebo. The purpose of the experiment was to measure the effect of caffeine administration upon the simultaneous brightness contrast function.

The resulting shape of the contrast function was similar to those obtained in previous experiments; i.e., enhancement occurred for inducing-luminance values below the test-field luminance, depression occurred for inducing luminances greater than test luminance. With caffeine administration the enhancement phenomenon seemingly disappears. However, depression still occurs and there does not seem to be much difference between the drug and nondrug state with respect to depression.

References

Davison, F. R. *Synopsis of materia medica, toxicology and pharmacology.* St. Louis: Mosby, 1949.

Diamond, A. L. Foveal simultaneous brightness contrast as a function of inducing- and test-field luminance. *J. exp. Psychol.*, 1953, **45,** 305–414.

Diamond, A. L. Theory of depression and enhancement in the brightness response. *Psychol. Rev.*, 1960, **67,** 168–199.

Diamond, A. L., Scheible, H., Schwartz, E., & Young, R. A comparison of psychophysical methods in the investigation of foveal simultaneous brightness contrast. *J. exp. Psychol.*, 1955, **50,** 171–174.

Granit, R. *Receptors and sensory perception.* New Haven: Yale Univer. Press, 1955.

Guilford, J. P. *Psychometric methods.* New York: McGraw-Hill, 1954.

Hartline, H. K., & Ratliff, F. Inhibitory interaction of receptor units in the eye of the *Limulus*. *J. gen. Physiol.*, 1957, **40,** 357–376.

Heinemann, E. G. Simultaneous brightness induction as a function of inducing- and test-field luminance. *J. exp. Psychol.*, 1955, **50,** 89–96.

Leibowitz, H., Mote, F. A., & Thurlow, W. R. Simultaneous contrast as a function of separation between test and inducing fields. *J. exp. Psychol.*, 1953, **46,** 453–456.

(Received November 20, 1959)

Journal of Experimental Psychology
1961, Vol. 61, No. 1, 23–29

RECOGNITION MEMORY FOR NOUNS AS A FUNCTION OF ABSTRACTNESS AND FREQUENCY [1]

ALOYSIA M. GORMAN [2]

University of Denver

Abstract words have been found harder to remember than concrete words whether learning is measured by recall (Stoke, 1929) or by recognition (Jampolsky, 1950). Little attention, however, has been given (*a*) to defining the universe from which the experimental words were obtained and securing representative sampling therefrom, (*b*) to partialing out the differential effect of total previous exposure to the word, as measured, for instance, by Thorndike-Lorge frequency counts, or (*c*) to the question whether the meanings of all nouns may be considered to lie on a psychological dimension of abstractness. Studies that apply scaling procedures to nouns in order to measure meaning or meaningfulness make little reference to an abstract-concrete dimension (e.g., Noble, 1952), although Underwood and Richardson (1956) take account of it as a possible variable. Osgood, Suci, and Tannenbaum (1957) have not identified an abstractness-concreteness dimension. However, such a dimension must refer primarily to a denotative (cognitive) aspect of meaning rather than an affective (connotative) aspect, and so may be excluded by their definition.

The present study was undertaken to provide an operational definition of abstractness in nouns as a dimension of cognitive meaning and to demonstrate the psychological usefulness of this definition by showing its relation to some accepted psychological variable. The human discriminative response was used to scale two points on the hypothesized dimension. The effect on short-term recognition memory of the variable so scaled was experimentally demonstrated, under the hypothesis that the relation of the scale to memory is such that nouns symbolizing concepts of higher degrees of abstractness tend to be less well retained than those symbolizing concepts of lower degrees of abstractness.

The operational definition has been directed exclusively toward the abstract-concrete dimension as distinguished from a possibly different general-specific dimension, and since it was desired to refer to the meaning of individual words independent of specific utterances, only pure nouns have been scaled. The experimental demonstration differs from the usual recognition learning experiment in its specific emphasis on cognitive meaning.

Construction of a Two-Point Abstractness Scale

Method

A two-copy list of 1791 nouns was prepared by including every single-word entry

[1] This paper is based on a dissertation presented to the faculty of the Graduate College, University of Denver, in partial fulfillment of the requirements for the degree of Doctor of Philosophy. The writer gratefully acknowledges the supervision and guidance of Joel E. Greene. She also wishes to express appreciation for the advice and assistance of the other members of her committee and for the aid of Edith L. Schnell, who acted as a judge in the construction of the operational definition.

[2] Now at Richmond Professional Institute of the College of William and Mary, Richmond, Virginia.

in Thorndike and Lorge (1944) that (*a*) had a G-frequency index higher than 6; (*b*) was not capitalized, abbreviated, contracted, or hyphenated; (*c*) was not expressly indeterminate as to frequency; (*d*) was not a plural of an entry that could be retained in the singular; and (*e*) did not have a substantial nonnoun use.

After training themselves on other nouns having a G-frequency index of 1, two judges independently classified the 1791 nouns into two categories, representing two levels of abstractness. Judge I, an experienced teacher of grammar, composition, and rhetoric at the college and graduate levels, worked from the *New World Dictionary* (1953); Judge II, the *E*, worked from the *American College Dictionary* (Barnhart, 1948). Only meanings found in one of these college dictionaries were considered; but each judge consulted other dictionaries when more information about such meanings was desired. To enable the judges to split the dimension at approximately the same point, the following instructions were used:

> The purpose is to distinguish two classes of nouns: (*a*) "concrete"—being those whose reference to objects, to material, to sources of sensation is relatively direct; (*b*) "abstract"—being those whose reference to objects, to material, to sources of sensation is relatively indirect. A word may be "abstract" *and* either general or specific, or "concrete" *and* either general or specific. Classify as "abstract" all nouns usually classed by grammarians and logicians as abstract in the sense opposed to concrete; also all nouns that are primarily names of measures, of processes, of kinds of persons characterized by reference to an a-sensory trait (e.g., *optimist*), and any others judged analogous to these. Classify as concrete the names of mythical animals or monsters, and all words judged analogous to these. Disregard any meanings that are judged to be "unfaded" metaphor (e.g., *gadfly*—a person who irritates others). Apply the same principle of reference to sight, hearing, taste, smell, and somesthesis.
>
> Assign every word to either the "concrete" or the "abstract" category. Add subscripts when necessary: *m* to indicate that while the word belongs predominantly to one category, some of its meanings belong to the other category, or *?* to indicate that assignment to the category chosen is felt to be uncertain.

Results

The judges agreed as to the category of 1713 or 95.6% of the 1791 nouns. They agreed as to both category and subscripts in 1444 cases. Of the 1444 nouns as to which the judges agreed completely, 723 were classified as abstract (with or without subscripts) and 721 as concrete (with or without subscripts). Both judges used the Abstract category without subscripts in 490 of these cases and the Concrete category without subscripts in 571.

The 1061 nouns classified without qualification by both judges as either abstract or concrete are considered to constitute an operational definition specifying two points on a psychological scale of abstractness.[2]

Influence of Abstractness on Recognition Memory

Method

Subjects.—The *S*s were 32 students from two sections of introductory psychology, selected and assigned to conditions by a random number procedure. Eight men and eight women were selected from each section, forming 4 administration groups—Section 1, men; Section 1, women; Section 2, men; Section 2, women.

Design.—In this experiment 240 different words were used; 192 were used as "background" words, 48 as "repeat" words. Each *S* saw all 240 words. Each *S* saw each of the "background" words once and each of the

[2] A list of the words of the operational definition (together with indexes for frequency, syllabication, and use of words in the experiment manipulating the scale as an independent variable) has been deposited with the American Documentation Institute. Order Document No. 6456 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress; Washington 25, D. C., remitting in advance $1.75 for microfilm or $2.50 for photocopies. Make checks payable to: Chief, Photoduplication Service, Library of Congress.

The author is also making available a limited number of copies of this material.

"repeat" words twice; i.e., no *S* received more than one trial on the same words.

The words were divided into 8 reading lists of 18 words each and 8 corresponding recognition lists of 18 words each. In each recognition list 6 words were repeated from the previous reading list and 12 were new words. Each combination of a reading list with its corresponding recognition list is herein called an observation. There was no repetition of words between any observation and any other observation.

To investigate the possible effect of word sequence and contiguity, each list of words was presented in one sequence to 16 of the *Ss* and in another sequence to the other 16. The sequences were determined by two series of shufflings (Shuffle 1 and Shuffle 2) of the 16 packs of cards from which the lists were prepared. Within each shuffling series, the 6 repeated words in an observation were shuffled first with the reading pack, then with the recognition pack, always under the restriction that no "repeat" word should be first or eighteenth on a list and that no more than two "repeat" words should come together in sequence.

All combinations of two levels of Abstractness and two levels of Frequency were administered to each of the 32 *Ss*; thus there were four treatment combinations: Abstract-High (AH), Abstract-Low (AL), Concrete-High (CH), and Concrete-Low (CL). There were two observations for each treatment combination, but for each *S* the two scores in a treatment combination were totaled and the mean taken as the single unit in each cell.

The levels of the Abstractness factor were defined, respectively, as the 490 abstract and 571 concrete nouns of the operational definition. The levels of the Frequency factor were defined as follows: Low—all nouns of the operational definition having Thorndike-Lorge Indexes 7 through 17; High—all nouns of the operational definition having indexes higher than 17.

The order of treatment combinations was determined by a "systematic randomization procedure" as discussed by Underwood (1949). This resulted in the following basic square: AH, AL, CH, CL; CH, AH, CL, AL; AL, CL, AH, CH; CL, CH, AL, AH. There were eight basic squares with four *Ss* in each. In each administration group there were two basic squares, one for each shuffle. Within each shuffle, there was one *S* in each administration group receiving a particular Order of treatment combinations, the four together forming an Order group. With shuffles disregarded, eight *Ss* formed an Order group.

This design, described by Lindquist (1953) under the heading Type VI, provides precise Within-*Ss* estimates for the effects of Abstractness and Frequency and for all interactions of Abstractness, Frequency, and Order, together with a Between-*Ss* estimate of the effect of Order.

Materials.—Preparatory to selecting the 240 experimental nouns from the 1061 nouns of the operational definition, each of the 1061 nouns was typed on a small card, and on the back of each card, indexes were entered for frequency, category, syllabication, and "ending" of the word. The "ending" index showed relative rarity of the "ending" and whether the ending was considered characteristic of the category, noncharacteristic, or characteristic of the opposite category.

Thirty different words were required for each observation: 6 "repeat" words and 24 "background" words. Hence, for each observation, five parallel sets of six words each were prepared, each set having two 1-syllable, two 2-syllable, and two 3-syllable words. No repetition of initial letter or of "ending" was allowed within a set, and lengths of syllables of within-set pairs were varied as much as possible. Words with rare, opposite, and noncharacteristic endings were preferred in order to secure further variation in the formal aspects of the words chosen. Each of the 20 sets within a frequency level was brought nearly to the same total frequency. The High level sets each had a total of 38 letters, the Low level a total of 39 letters.

The initial selection of words for the sets was made blindly by means of the indexes on the backs of the cards. After the blind balancing of indexes, the cards were turned over, and some further rearrangements and substitutions were necessary to bring the observations to a standard of variety in initial letters and endings. If obvious synonyms or pair-words, such as *uncle-aunt*, were noticed during this process, they were allocated to different observations, or, if this was impossible, substituted for. Substitutions were blindly drawn, and substitutions and rearrangements were allowed only according to rule. To control the residual possibility of intuitive selection, the five sets of each observation were assigned to conditions within the observation by a random number procedure carried out independently for each observation. Thus, in each observation, one set was assigned to Cond. 1, "repeat"; two sets were assigned to Cond. 2, "background" in the reading list; and two

TABLE 1

THE 240 WORDS USED IN THE EXPERIMENTAL DEMONSTRATION

Abstract Nouns

anecdote 7-224-AL21
appetite 24-224-AH12
approval 14-242-AL13
argument 48-224-AH13
behalf 18-24-AH12
birthday 37-53-AH23
bout 8-4-AL13
census 7-33-AL23
chaos 9-32-AL12
comedy 17-312-AL22
contempt 18-35-AH12
creed 14-5-AL21
crime A-5-AH23
debt A-4-AH23
decade 13-33-AL13
decision 46-224-AH12
defect 14-24-AL22
departure 23-234-AH21
dimension 10-234-AL22
disaster 23-323-AH23
doctrine 30-35-AH11
dynasty 9-232-AL11
emphasis 10-233-AL22
energy 41-222-AH12
enterprise 29-235-AH22
episode 7-214-AL12
errand 20-24-AH12
event A-14-AH11
existence 45-234-AH13
expense A-25-AH22
fact AA-4-AH23
famine 15-33-AL21
fate A-4-AH12
feat 13-4-AL13
fiction 14-34-AL11
fortune A-34-AH13
glee 9-4-AL22
grandeur 9-44-AL22
gratitude 22-414-AH11
grief 45-5-AH22
guilt 14-5-AL23
hatred 20-24-AH13
headache 10-44-AL12
hour AA-4-AH12
idleness 9-134-AL13
illness 22-34-AH22
income 46-24-AH23
injustice 12-234-AL21
intellect 8-234-AL12
language A-35-AH22
law AA-3-AH13
leadership 18-424-AH22
loss AA-4-AH21
magnitude 7-324-AL13
measurement 11-434-AL22
membership 13-334-AL23
method AA-42-AH13
miracle 24-313-AH22
mirth 16-5-AL12
mode 23-4-AH23
mood 27-4-AH11
morrow 7-33-AL11
myth 8-4-AL23
nonsense 19-35-AH13
oath 18-4-AH13
occurrence 9-235-AL13
odds 10-4-AL13
opinion AA-133-AH23
origin 28-213-AH21
outcome 8-34-AL12
ownership 9-324-AL11
pang 12-4-AL22
pastime 10-34-AL23
peace AA-5-AH11
percentage 13-343-AL23
phase 15-5-AL13
pint 14-4-AL11
plea 10-4-AL12
poetry 26-222-AH22
politics 40-314-AH13
pomp 12-4-AL21
principle A-423-AH12
procedure 11-324-AL12
proposal 20-332-AH13
pursuit 21-34-AH23
ratio 11-23-AL22
rebellion 14-334-AL12
response 18-26-AH21
revenue 19-313-AH11
revival 9-232-AL23
rhythm 7-6-AL12
rite 10-4-AL23
satire 7-33-AL13
secrecy 7-232-AL23
sentiment 29-324-AH23
series A-24-AH21
skill 39-5-AH13
soul AA-4-AH22
speech AA-6-AH12
stead 9-5-AL22
strife 17-6-AL11
tactics 7-34-AL22
tale A-4-AH13
tendency 27-422-AH23
tennis 18-33-AH12
threat 17-6-AL23
tidings 12-25-AL13
ton 49-3-AH22
topic 19-32-AH22
trait 7-5-AL22
treason 15-43-AL13
usage 7-23-AL12
valor 8-32-AL23
verdict 11-34-AL21
warfare 14-34-AL23
week AA-4-AH22
width 28-5-AH12
wisdom 49-33-AH23
wrath 21-5-AH21
zeal 15-4-AL12

Concrete Nouns

actress 13-25-CL11
admiral 20-223-CH21
alley 13-23-CL13
ambulance 8-225-CL22
ankle 21-23-CH12
anvil 7-23-CL23
apartment 49-144-CH13
archbishop 15-442-CL13
attorney 23-233-CH13
bacon 12-23-CL13
barn 45-4-CH22
basement 8-44-CL22
bedroom 35-34-CH21
beer 23-4-CH13
bulb 14-4-CL12
bullet 22-33-CH23
bungalow 8-323-CL21
butterfly 22-333-CH12
calcium 9-322-CL22
calf 14-4-CL13
candidate 45-324-CH12
canyon 7-33-CL21
car AA-3-CH22
chaplain 8-44-CL22
cigarette 22-315-CH11
cliff 38-5-CH13
coffee A-33-CH12
colt 21-4-CH23
corpse 9-6-CL23
desk A-4-CH11
disk 8-4-CL11
engine A-24-CH13
factory A-322-CH12
farmer AA-42-CH13
fern 13-4-CL23
filament 7-314-CL13
foliage 12-223-CL23
furnace 29-34-CH11
gelatin 7-313-CL12
gentleman AA-333-CH21
grape 34-5-CH12
grove 31-5-CH22
hermit 16-33-CL22
hillside 20-44-CH12
hospital A-323-CH13
hotel A-23-CH23
isthmus 13-43-CL23
knob 7-4-CL22
lake AA-4-CH11
lane 32-4-CH13
lantern 19-34-CH23
leaflet 7-43-CL13
ledge 14-5-CL12
lieutenant 33-433-CH23
luggage 10-34-CL12
lung 15-4-CL21
mare 14-4-CL22
marsh 16-5-CL12
microscope 9-235-CL12
millionaire 14-344-CL23
monarch 20-34-CH12
monk 20-4-CH13
moth 9-4-CL13
mule 29-4-CH23
museum 38-222-CH22
mustache 7-35-CL21
necklace 9-44-CL23
newspaper AA-423-CH23
nymph 13-5-CL13
ocean AA-14-CH13
onion 25-23-CH21
orchestra 10-243-CL21
ore 18-3-CH12
overalls 7-134-CL11
palace A-33-CH22
physician 43-324-CH23
pineapple 15-423-CL11
policeman 22-243-CH11
pond 30-4-CH23
poplar 8-33-CL12
porch A-5-CH21
porcupine 7-324-CL22
potato A-222-CH22
prairie 25-43-CH13
restaurant 23-334-CH13
revolver 9-242-CL13
salad 28-32-CH11
seashore 9-35-CL12
senator A-313-CH23
shawl 9-5-CL13
sidewalk 18-44-CH22
skull 12-5-CL11
steak 14-5-CL23
strawberry 19-532-CH22
street AA-6-CH12
tavern 12-33-CL11
teacher AA-52-CH23
thief 28-5-CH21
tomato 11-222-CL12
tray 17-4-CL12
trophy 8-33-CL22
trousers 21-44-CH22
trout 17-5-CL22
tulip 8-23-CL23
uncle AA-23-CH22
underbrush 8-235-CL12
valentine 7-324-CL13
vase 12-4-CL23
vine 38-4-CH23
vinegar 12-313-CL23
volcano 14-322-CL22
waist 33-5-CH22
walnut 11-33-CL12
warrior 37-322-CH12
wasp 13-4-CL21
waterfall 10-234-CL23
wife AA-4-CH12
wilderness 34-334-CH22
windmill 11-44-CL13
wrist 17-5-CL22

sets were assigned to Cond. 3, "background" in the recognition list.

The 240 words thus selected and assigned appear in Table 1, each word being followed by three indexes. The first is its Thorndike-Lorge G-frequency index; the second is a syllabication key; the third shows in what treatment combination, observation, and set-condition the word was used. For example, 7-224-AL21 following *anecdote* shows that the word (*a*) has a G-frequency index of 7, (*b*) has three syllables, the first and second having two letters and the third four letters, and (*c*) was used in the Abstract-Low treatment combination, in the second of the two observations in that treatment combination, and in the first of the set-conditions, i.e., it was repeated.

For administration purposes, each list was triple spaced, one word to a line, on a half sheet of paper, with a row of stars at the top and bottom of the list. The lists were reproduced commercially, and were stapled into booklets in the order of treatment combinations determined by the design. The order of observations within a treatment combination was individually randomized for each treatment combination and each *S*.

A covering sheet with a window strip was attached to each list on the "sales book" principle in such a way that only one word at a time could be seen by *S*. Each reading list had a blue covering sheet, each recognition list a salmon covering sheet. Each covering sheet was the same width as the list sheet and about twice as long. It was attached at the bottom of the list sheet and turned up over the list sheet in such a way that the window strip came at the top and a loose flap had to be pulled out from behind the window strip before the row of stars at the top of the list was exposed. The *S* used one of the tabs of the window strip to pull the covering sheet down the list and to hold the strip steady when he reached the bottom row of stars. He then creased the loose fold of the covering sheet which had appeared next to his body as the covering sheet was pulled down, and threw the covering sheet (the loose flap of which had been drawn down over and was still covering the list) back by the creased fold. He then pulled the loose flap of the next covering sheet out from under its window strip to expose the top row of stars and proceeded as before.

Practice booklets were also provided, made up exactly like two observations of the experimental booklet except that sign-groups (e.g., ::::::) were used instead of words, and the number of "repeat" sign-groups was 5 for one "observation" and 7 for the other.

Procedure.—The four administration groups were run separately on the same day. After a demonstration of the manipulation of the booklet, *Ss* were given a practice period under the same conditions as in the experimental period. They then went through the eight observations of the experimental booklet continuously without rest periods. They were allowed 1 sec. for each word, were monitored, and were paced with an automatic buzzer set at the rate of 1 buzz per sec. They were told that *some* of the words on each reading list would be repeated on the corresponding recognition list, and that they were to mark the words that were repeated. They were told to *read* the words rather than try to memorize them. The key instructions were as follows:

> Set yourself to read each word silently for *meaning*. Pay attention to the meaning of each word individually; don't try to associate it with other words on the list.
>
> Try not to mark any words that are not repeated. But if you make a mistake, do not erase or change, go right on.
>
> Remember: Concentrate on the meaning of each word individually. Give equal attention to each word.

The key instructions to *Ss* were designed to produce an internal reading set as nearly as possible equivalent to: "Relate each symbol to that which it symbolizes."

Besides the 1 sec. for each word, time was allowed for manipulating the covering sheets, so that the total time allowed for the experimental booklet was 6 min. 40 sec. All but five *Ss* completed the booklet in the allowed time, the slowest *S* finishing in 6 min. 50 sec.

Scores were obtained for each recognition list by weighting each correct mark +1 and each incorrect mark −1, and adding algebraically. For each *S*, the mean of the combined scores of the two observations in a treatment combination was the unit of analysis.

Results

The means and *SD*s of the treatment combination scores of all *Ss* are shown in Table 2. The following analyses were performed and showed no differences or interactions significant at the .05 level: (*a*) a χ^2 test for difference in the median times of the administration groups; (*b*) a

TABLE 2

MEANS AND *SD*s OF TREATMENT-COMBINATION SCORES OF ALL *Ss*

Measure	Treatment Combination			
	AH	AL	CH	CL
Mean	5.09	6.94	7.09	8.47
SD	2.74	2.37	2.43	2.65

Sex × Section × Shuffle analysis of the total scores; (*c*) a Sex × Section analysis of difference scores (Abstract − Concrete).

Table 3 shows the final analysis for all groups combined. The effects of both Abstractness and Frequency are highly significant ($P < .001$). There is no significant effect of Order as a Between-*Ss* variable, but the Within-*Ss* interaction of Abstractness and Order is significant at the .05 level. There are no other significant interactions.

DISCUSSION

The results have shown that the two-point scale of abstractness has adequate reliability and that the dimension of abstractness defined by this scale has an effect in the expected direction on short-term recognition memory. Some evidence was found that the degree of the abstractness effect within an *S* is different for different orders of the treatment combinations. However, since no Between-*Ss* effect of Order was found, it appears that each particular order may facilitate memory in some *Ss*, inhibit it in others.

The operational definition of the present study thus appears to furnish some basis for control of the abstractness variable in studies of learning and of meaning using English words. The frequency range from which the two-point scale was obtained covers approximately the 7600 most frequent English words according to the Thorndike-Lorge index. It is recommended by Thorndike-Lorge that words within this range be taught by the end of the eighth grade; beyond this point the recommendations are much less specific. Most studies of meaning include much rarer words than these, and the operational definition of abstractness could, of course, be extended to much rarer words. However, the range for which the effect on recognition memory may be validly demonstrated is restricted by the requirement that all words used should be familiar to all of the *Ss*. It is probable that when the denotative aspect (What does it mean?) is being stressed, the appropriate frequency range is more restricted than in the connotatively oriented study (What does it mean to you?).

Frequency was found to be a pertinent variable producing a highly significant effect even though all the words used may be supposed to be familiar to all the *Ss*. As previously indicated, the studies of Stokes (1929) and Jampolsky (1950) and the present study are consistent in finding that abstract words are less well remembered than concrete words. However, it appears from the published reports that in the two earlier studies cited there was no investigation or control of frequency beyond that of using only reasonably familiar words.

TABLE 3

ANALYSIS OF VARIANCE OF SCORES OF ALL *Ss*

Source	*df*	*MS*	*F*
Between-*Ss*	31		
Order	3	.98	
$error_b$	28	3.68	
Within-*Ss*	96		
Abstractness	1	24.94	25.19**
Frequency	1	20.72	23.02**
A × F	1	.44	
A × O	3	3.41	3.44*
F × O	3	2.44	2.71
A × F × O	3	.71	
$error_w$	84		
e_1	28	.99	
e_2	28	.90	
e_3	28	.83	
Total	127		

* $P \leq .05$.
** $P < .001$.

Hence, the present study indicates (*a*) that the results of earlier studies as to the "recall or recognition value" of abstract and concrete words may have been open to interpretation as wholly or partially due to uncontrolled differences in frequency, and (*b*) that when frequency is controlled a difference is still found in the "recognition value" of abstract and concrete nouns.

In the results of the present study, Frequency was found to be independent of Abstractness, and no significant interaction was found between Frequency and Order of treatment combinations. The direction of the frequency effect was unusual, very frequent nouns being less well remembered than somewhat rarer nouns. This negative relationship between frequency and recognition may be plausibly accounted for by the hypothesis that the high-frequency words were more subject to interference effects because of their high frequency of occurrence with other words in the common language. It should be noted, however, that no differences were found connected with sequence or contiguity of words on the lists. The negative relationship may, on the other hand, be specific to one or more of the techniques employed, i.e., using only pure nouns, only one-trial learning, only 1-sec. exposure, and instructing *Ss* to read for meaning rather than memorize.

Summary

An operational definition of abstractness in nouns was constructed by using the human discriminative response to identify two points on a scale of abstractness. This scale, consisting of 490 "abstract" and 571 "concrete" nouns, was found to have adequate reliability. When the scale was manipulated as an independent variable, the effect of abstractness on short-term recognition memory was highly significant, "abstract" nouns being less well remembered than "concrete" nouns. Frequency was found to be a pertinent variable, independent of abstractness, very frequent nouns being less well remembered than somewhat rarer nouns.

References

Barnhart, C. L. (Ed.) *The American college dictionary.* (Text ed.) New York: Harper, 1948.

Jampolsky, M. Étude de quelques épreuves de reconnaissance. *Annee psychol.*, 1950, **49**, 63–97.

Lindquist, E. F. *Design and analysis of experiments in psychology and education.* Boston: Houghton Mifflin, 1953.

Noble, C. E. An analysis of meaning. *Psychol. Rev.*, 1952, **59**, 421–430.

Osgood, C. E., Suci, G. J., & Tannenbaum, P. H. *The measurement of meaning.* Urbana: Univer. Illinois Press, 1957.

Stoke, S. M. Memory for onomatopes. *J. genet. Psychol.*, 1929, **36**, 594–596.

Thorndike, E. L., & Lorge, I. *The teacher's word book of 30,000 words.* New York: Bureau of Publications, Teachers Coll., Columbia Univer., 1944.

Underwood, B. J. *Experimental psychology.* New York: Appleton-Century-Crofts, 1949.

Underwood, B. J., & Richardson, J. Some verbal materials for the study of concept formation. *Psychol. Bull.*, 1956, **53**, 84–95.

Webster's new world dictionary of the American language. (College ed.) Cleveland: World Publishing, 1953.

(Received December 1, 1959)

Journal of Experimental Psychology
1961, Vol. 61, No. 1, 30–39

THE JUDGMENT OF SIZE, CONTRAST, AND SHARPNESS OF LETTER FORMS [1]

WILLIAM C. HOWELL AND CONRAD L. KRAFT [2]

Ohio State University

A considerable portion of man's normal visual activity is devoted to identifying complex figures, particularly alphabetical symbols, under a variety of viewing conditions. Although the literature abounds with studies of legibility and the variables related to it (Berger, 1944; Crook, Harker, Hoffman, & Kennedy, 1950; Lansdell, 1954; Schapiro, 1952; Spragg & Rock, 1952; Tinker, 1928), relatively little research has been directed toward quantifying the perceptual continua involved in legibility. To be sure, psychophysical research has produced data regarding brightness discrimination and visual acuity, but the perceptual differences between a clear and a blurred letter, or between a large and a small one, for example, have not been subjected to quantification. The major purpose of this investigation, therefore, was to quantify psychophysically three variables involved in figure recognition: image size, blur, and contrast. The importance of these stimulus dimensions for figure recognition has been described recently by Howell and Kraft (1959). Although letters of the alphabet were used as stimuli in most of the present research, it is suggested that the findings have implications for the psychophysical description of complex figures in general.

A second purpose of this research was purely methodological. The recent development of techniques for scaling directly perceived magnitudes has led to the redetermination of a number of classical psychophysical functions (Stevens, 1956; Stevens & Galanter, 1958). Briefly, these techniques are based upon the ability of *S* to judge magnitude relative to a standard value and to report these judgments in terms of the number system. In contrast to the classical psychophysical techniques, many of which result in scales with ordinal properties, the new methods are purported to yield ratio scales of perceived magnitude (Stevens, 1957, 1958). In addition to this superiority, the magnitude-estimation techniques have the practical advantage of reducing significantly the time and labor necessary to obtain reliable psychophysical functions. To date, the application of these methods has been restricted primarily to classical problems involving simple stimuli and basic stimulus attributes. The intent of this study was to examine the efficiency of these methods in dealing with more complex stimulus materials. More specifically, answers were sought to the following questions: (*a*) Is *S* able to judge and assign numerals to magnitude in a consistent fashion for relatively complex dimensions such as size and blur? (*b*) Is *S* able to abstract these dimensions from

[1] This research was conducted in the Laboratory of Aviation Psychology of Ohio State University and was supported in part by the United States Air Force under Contract No. AF 33(616)-3612 with the Ohio State University Research Foundation, monitored by the Aero Space Medical Laboratory, Wright Air Development Division. Permission is granted for reproduction, publication, use, and disposal in whole or in part by or for the United States Government.

[2] Now with Boeing Airplane Company, Seattle, Washington.

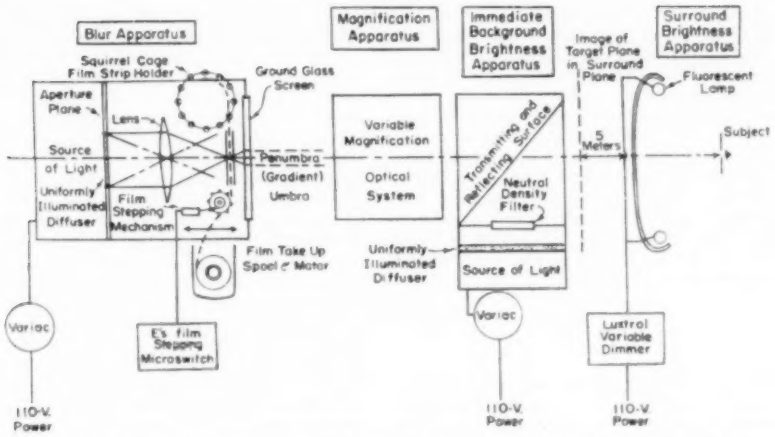

FIG. 1. Schematic drawing of the principal elements of the experimental apparatus.

complex, meaningful stimuli such as letters of the alphabet? (*c*) Do the functions obtained for these dimensions correspond to the functions obtained by others for basic dimensions involving less complex stimuli? To the extent that affirmative answers can be given to these questions, the data may be considered powerful evidence for the generality of the direct magnitude-estimation methods.

Three experimental investigations were undertaken in order to meet the above objectives. In each, the method of magnitude estimation described by Stevens (1956) was used to scale the stimulus attributes. The first study was concerned with the scaling of apparent size, sharpness,[3] and contrast individually under optimal conditions of the other two dimensions. The second and third studies were aimed at determining the extent to which judgments along the individual continua are independent of the level of the other two stimulus dimensions.

[3] The term *sharpness* is used whenever judgment is involved; *blur* refers to the physical dimension.

GENERAL METHOD

Apparatus.—The device used to project the stimuli and to manipulate their size, blur, and contrast levels was loaned to the authors by Glenn A. Fry and Jay M. Enoch and the Mapping and Charting Research Laboratory of the Ohio State University. This apparatus was designed originally for use in photo-interpretation research, and the details of its construction are reported elsewhere (Fry & Enoch, 1957). Figure 1 shows the major components of the apparatus together with the modifications introduced for the purpose of this study. Light from a standard projection lamp was passed through a diffuser and collimated by a lens which focused on a ground glass screen. The stimuli, which were introduced just prior to this point on 35-mm. filmstrips, were projected on the ground glass screen. The filmstrips were housed in a portion of a Soundview 35-mm. projector which included the storage, holding, and stepping mechanisms. The whole filmstrip assembly was mounted on a movable platform which permitted the distance between the ground glass screen and film to be varied. By manipulating this distance, it was possible to vary the width of the penumbra or gradient of all light-dark transitions projected on the screen. The principle involved in this method of producing blur has been described in detail by Fry and Enoch (1959). Briefly, the introduction of a target between the lens and the ground glass screen creates a shadow on the screen, the

width of which shadow increases with the distance (d) of the target from the screen. The distribution of illuminance across each such light-dark transition is sigmoid in shape regardless of target orientation owing to a circular aperture in the plane of the diffuser. The *shape* of this distribution remains constant despite movement of the target: only the *width* of the distribution changes. Thus, the amount of blur may be specified with reference to the light distribution, a convenient index (ϕ) being the ratio of the distribution maximum (E_{max}) to the slope of the distribution at its midpoint, or

$$\phi = \frac{E_{max}}{S}$$

This index accounts for the near-linear portion of any blur gradient. For the particular arrangement of components used, the value of ϕ was directly proportional to d, the distance between the target (film) and the screen, the function being

$$\phi = .0825\ d$$

Thus, it was possible to manipulate blur (or sharpness) through the manual adjustment of d, and to specify the resulting amount of blur in distance covered by the light distribution at the screen. Actually, this index was modified for use in the present work in an attempt to achieve a more meaningful specification of blur as a characteristic of letter stimuli. The modification consisted of dividing all ϕ values by the stroke width of the letters. The assumption was that the blurredness of a letter is more adequately reflected in the width of transition *relative* to the width of constant illuminance than in the width of transition alone.

The image on the ground glass screen was viewed through a zoomar-type magnification system and a half-silvered mirror illuminated from above by a constant level of diffused light. The former provided a convenient method for manipulating the angular size of the stimuli, and the latter served as the background illuminance for the stimuli on the basis of which contrast specifications could be made. Figure illuminance was manipulated by means of a continuous voltage control (Variac) inserted between the 110-v. power line and the projector lamp. Both the voltage control and the magnification system were manually operated.

In order to control accommodation, a collimating lens was introduced between the eye and the rest of the optical system which effected accommodation for infinity. The *S* viewed the stimulus through a 2-mm. artificial pupil, the image appearing in the center of a limitless, uniformly illuminated field. The illuminance of this background (3.5 ft-c) was such that visual adaptation could be considered complete within seconds after exposure in the apparatus.

The size, blur, and contrast levels of the stimuli employed in this apparatus were specified as follows: (*a*) size, in minutes of visual angle of letter height, except as otherwise noted; (*b*) blur, in units based on the ratio of light-dark transition width to stroke width; and (*c*) contrast, in the ratio of figure-minus-background to background illuminance. This method of specifying contrast is one of three in common use. Since the others involve multiplication by a factor of 100 to yield a percentage measure, they would have resulted in extremely cumbersome values in the present case; hence, the use of the brightness ratio.

Scaling procedures.—The method of magnitude estimation (Stevens, 1956, 1957) was used for scaling all dimensions investigated in the three studies. This method was applied in the following manner. The *S* was presented with successive pairs of stimuli, the first in each pair being a constant value (the standard or modulus) and the second being drawn randomly from a population of 10 values spaced evenly along the stimulus dimension (e.g., size). The standard was designated by the number "10" and *S* was required to estimate the value of the variable on the basis of this standard. His response consisted of assigning a number to the variable in accordance with its perceived magnitude (e.g., if the variable appeared twice as large as the standard, it was called "20"). The standard was presented for 2 sec. and the variable for the length of time until a response occurred (usually less than 1 sec.). Approximately 1 sec. separated the standard and the variable, and 5 sec. elapsed between pairs. Forty judgments, four of each variable, were made in an experimental session with a 5-min. rest period separating each series of 20. No more than two sessions were undertaken in a single day.

Experiment I

Method

Subjects.—Sixteen college students served as *Ss*, judging size, sharpness, and contrast in a random sequence on separate sessions. Half of these *Ss* had served in previous

investigations using this apparatus, whereas the others were completely unfamiliar with the experimental arrangement.

Stimuli.—Four letters of the alphabet, chosen so as to maximize differences in width, shape, area, and ease of recognition were used as stimulus figures. These letters were I, O, H, and F. They were arranged on 35-mm. filmstrips in four quasi-random sequences of 80 stimuli each for presentation in the projecting arrangement described above. The four sequences differed in the proportion of times the standard and variable conditions appeared using different letters; the proportions being .00, .50, .75, and 1.0. This procedure was adopted to permit an estimate of the effect of standard-variable figure heterogeneity on dimensional judgments.

Procedure.—In this study, each of the three dimensions was scaled with the other two at their optimal values (e.g., contrast was scaled with sharpness and size at their maxima). The presentation sequences (i.e., filmstrips) were assigned so as to allow four *Ss* to judge each dimension under one of the four sequences. Since each sequence represented a different proportion of standard-variable figure heterogeneity (see above), this resulted in 16 judgments for each magnitude level under each degree of heterogeneity.

The 10 stimulus levels used for each dimension are given in Table 1. In each case the upper limit was apparatus-imposed and the lower limit was an approximation of the letter recognition threshold (75%). For each condition half of the *Ss* judged against a standard set at the fifth magnitude level and half against one at the sixth level. In both cases the standard was designated "10." This procedure was intended to test the validity of an assumption inherent in the magnitude estimation technique: namely, that *S* uses numbers in the same manner to describe events at all points on the continuum. The implications of this assumption are discussed more fully in a later section.

TABLE 1

PHYSICAL SPECIFICATION OF THE STIMULUS LEVELS USED FOR EACH DIMENSION IN EXP. I AND II

Level	Dimensions		
	Size (Min. of Visual Angle)	Blur $\left(\frac{\text{Distribution}}{\text{Stroke Width}}\right)$	Contrast $\left(\frac{B_1 - B_2}{B_2}\right)$
1	4.73	0.00	1.43
2	8.20	0.55	4.14
3	11.25	1.04	8.14
4	15.00	1.54	12.14
5	18.25	2.09	16.14
6	21.60	2.58	20.71
7	24.60	3.13	25.29
8	27.00	3.63	29.39
9	31.00	4.18	33.29
10	36.79	4.67	37.57

TABLE 2

ANALYSES OF VARIANCE PERFORMED ON THE DATA OF EXP. I

Source	*df*	Size		Sharpness		Contrast	
		MS	*F*	*MS*	*F*	*MS*	*F*
Between-*Ss*	15						
S-V heterogeneity (H)	3	24.88	5.7*	5.88	2.2	5.37	1.4
Standard magnitude (S)	1	149.92	34.4***	114.89	42.2***	45.06	12.0***
S × H	3	6.62	1.5	5.01	1.8	.95	.2
Error (b)	8	4.36		2.72		3.76	
Within-*Ss*	144						
Variable magnitude (V)	9	528.84	319.1***	468.28	142.3***	318.29	127.5***
V × H	27	3.80	2.8***	2.53	.8	1.97	.8
V × S	9	5.46	4.0***	17.64	5.4***	2.21	.9
V × H × S	27	2.39	1.8*	3.82	1.2	1.76	.7
Error (w)	72	1.35				2.50	
Total	159						

* $P < .05$.
** $P < .01$.
*** $P < .001$.

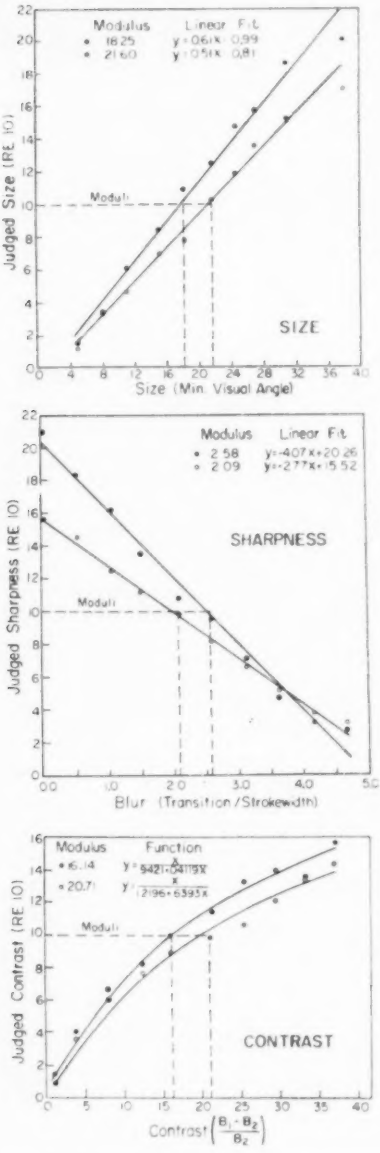

FIG. 2. A plot of the mean estimates of magnitude for the three dimensions of Exp. I.

Results

The influence of standard magnitude (S), standard-variable figure heterogeneity (H), and variable magnitude (V) on judged magnitude was tested for significance by an analysis of variance. Variable magnitude (V) was included in the design primarily to permit an evaluation of the V × S interaction (it being a near certainty that the effect of variable magnitude would prove significant). A separate analysis was performed for each of the three dimensions studied, and the results of these analyses are given in Table 2. It may be seen that in addition to variable magnitude (V), only standard magnitude (S) of the main effects exerted a significant ($P < .01$) influence on judgments for all three dimensions. Standard-variable figure heterogeneity (H) reached significance ($P < .05$) for only the size dimension. Significant V × S interactions ($P < .001$) were obtained for size and sharpness, as was the V × H interaction for size.

To permit an interpretation of these findings, the psychophysical functions were plotted for each dimension using arithmetic mean judgment values (Fig. 2). Arithmetic means were used in preference to the customary median values because the problem of extreme judgments did not arise (see Stevens, 1956). In fact, functions based on median values were also plotted and were found to be nearly identical to the mean curves shown. In determining the means, the data from all four levels of figure heterogeneity (H) were combined; only the S and V variables are shown in the figure. For both size and sharpness the data are described adequately by linear equations. For contrast, hyperbolic equations provide a more adequate fit.

The specific equation for each function is listed in the figure. It will be noted in Fig. 2 that for size and sharpness, the functions based upon different moduli have quite dissimilar slopes, whereas those for contrast do not exhibit such large slope differences. This finding is borne out in Table 2 by the significant V × S interactions for size and sharpness and the insignificant V × S interaction for contrast. The method of least squares was used to fit all curves. Indications of the variability of judgments were omitted from the figures to avoid unnecessary clutter. For size, the σ's ranged from .5 magnitude units at the smallest size to 4.5 at the largest size. For sharpness, the extremes of variability were .6 and 5.1 magnitude units, and for contrast, the values were .6 and 2.7 magnitude units.

It may be noted that the judgments of sharpness are inversely related to the corresponding physical values (blur units). The reason for expressing the latter as blur rather than sharpness is the fact that blur is quantifiable and possesses a zero point, whereas sharpness cannot be so easily specified. Sharpness, on the other hand, was chosen as the variable to be judged rather than blur to avoid confusion on the part of *S*: for all dimensions higher numbers referred to "better" conditions.

Experiment II

Method

Subjects.—Five students enrolled in the Ohio State University School of Optometry served as *S*s in this experiment. All five were familiar with the experimental situation.

Stimuli.—The stimuli used were the same as those described in Exp. I. Only one level of standard-variable figure heterogeneity (.50) was used.

Procedure.—The question asked in this study was: Are judgments along one dimension independent of the values of the other two? To answer this question, each dimension (size, sharpness, contrast) was scaled with the other two at 75% recognition threshold values. These values were for size, 4.73′; for contrast, 1.43%; and for sharpness, 4.67 blur units (ratio of transition width to stroke width). All scaling procedures were identical to those employed in Exp. I. A comparison of the functions obtained under these *minimum* conditions with those obtained under *maximum* conditions (Exp. I) was expected to reveal the maximum extent of any interaction among dimensions.

Results

The functions obtained under these minimal conditions are shown in Fig. 3. To aid in comparing these results with those of Exp. I (maximum conditions), best-fit lines for the latter functions were fitted to the present data. It is apparent that for size and contrast, the fits are quite good. In the former case, an extremely good fit is obtained, although in the latter, the data points appear to fall slightly below the best-fit line for maximum conditions and variability is rather large. For sharpness the fit is quite poor. In fact, variability is so great under this condition that any conclusion would seem tenuous. It should be noted, in this regard, that any of the above physical dimensions could be reduced sufficiently to make all other judgments impossible. This condition would be expected to occur near *absolute* threshold values rather than letter *recognition* thresholds as used here. Since sharpness judgments are disrupted by lowering size and contrast to recognition threshold levels, and since neither size nor contrast judgments are disturbed by such reductions, it must be concluded that perceived sharpness is relatively dependent upon size and contrast values (at least at these low levels). As for perceived size

and contrast, on the other hand, these dimensions appear independent of the others, insofar as the others are manipulated together. The effect upon these perceptual continua of the other dimensions manipulated *independently* is approached in the succeeding experiment.

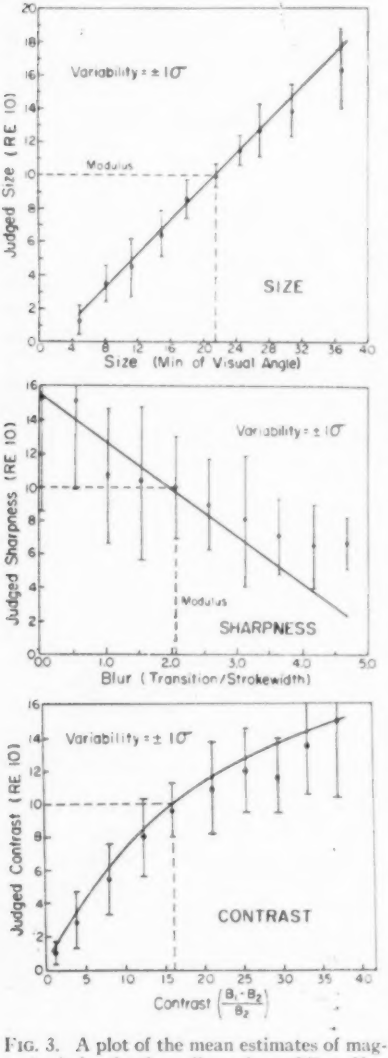

FIG. 3. A plot of the mean estimates of magnitude for the three dimensions of Exp. II.

EXPERIMENT III

Method

Subjects.—Four of the five *S*s who participated in Exp. II also served in this study.

Stimuli.—A rectangular figure resembling the letter "I" was used throughout this study. Its width was varied systematically on the film strips between .25′ and 16.00′ visual angle, and its height remained constant at 36′. This figure was used in preference to other letters of the alphabet to permit the judgment of exceptionally narrow sizes (near minimum visible threshold). Ten widths (.25′, .50′, 1′, 2′, 4′, 6′, 8′, 10′, 13′, and 16′ visual angle) were used, these being spaced more closely at the lower end of the scale than at the upper end. Further, the standard width was placed at a point near the lower end of the scale (2.0′). The reason for this arrangement was to provide ample opportunity for any influence of contrast and blur on size (in this case, width) judgments to appear. It was expected that if such interactions occurred, they would be most prominent at the near-threshold sizes.

Procedure.—Having established the points that (*a*) suprathreshold size, sharpness, and contrast can be scaled reliably using the method of magnitude estimation and (*b*) the reduction of two of the three dimensions *at once* has little effect on judgments of size and contrast and an appreciable but undetermined effect on judgments of sharpness, a study was designed to determine the influence of varying two dimensions (sharpness and contrast) *individually* on judgment of a third (size). It was suspected that blur might tend to enlarge the apparent size by spreading out the stimulus energy, whereas reduced contrast would tend to shrink apparent size by reducing the over-all energy projected to the eye. That reducing both variables *at once* had no influence on size judgments in Exp. II might then be attributed to a mutual cancellation effect.

The 10 size steps were judged under three conditions: (*a*) a blur-free, 37.29% contrast condition (high sharpness, high contrast), (*b*) a 4.67-unit blur, 37.29% contrast condition (low sharpness, high contrast), and (*c*) a blur-free 4.43% contrast condition (high sharpness, low contrast). In all cases, the standard figure with which the judged stimulus was compared was blur-free and of maxi-

mum contrast. A total of 20 judgments was made of each stimulus under each condition. In all other respects, the procedure used in this experiment was identical to that employed in the previous two studies.

Results

The functions obtained are shown in Fig. 4. Here it may be seen that the hypothesis is substantiated: the reduced sharpness condition resulted in larger size judgments and the reduced contrast condition produced lower size judgments than the control condition. An analysis of variance performed on the data indicated that this effect was statistically significant at $P < .01$. Variability under both reduced conditions was similar to that for the nonreduced condition. In fact, the over-all variability values obtained in this experiment were definitely of the same order of magnitude as those obtained in Exp. I.

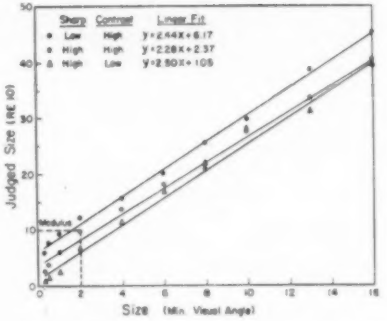

FIG. 4. A plot of the mean estimates of magnitude for size obtained under three combinations of sharpness and contrast values for Exp. III.

DISCUSSION

In all three experiments it is apparent that the method of magnitude estimation yielded systematic and reliable scales for each of the variables investigated. Only the judgments of sharpness under minimal size and contrast conditions appear to include serious amounts of variability, and this may be attributed to difficulty of judgments rather than to any inadequacies of the scaling technique. The present success of this method is significant for reasons pointed out earlier. First, it demonstrates that *S* is able to judge consistently the magnitude of relatively complex dimensions such as size and blur. Second, it suggests that *S* is capable of abstracting the dimensions for judgment from complex, meaningful stimuli without serious distortion of the resulting functions. This was clearly demonstrated by the consistency of judgments despite the use of widely differing stimulus figures, and by the general ineffectiveness of standard-variable heterogeneity to influence judgments consistently. Only for size did this variable reach marginal significance ($P < .05$). Even here, a close examination of the pattern of differences failed to reveal any consistent standard-variable heterogeneity effect: one group (75% heterogeneity, 18.25′ standard) appeared to be primarily responsible for the effect, and its judgments were considerably above *all* others including the two 100% heterogeneity groups.

A third indication of the success of the method used in this work relates to the primary objective of the research: the determination of the form of functions relating physical size, blur, and contrast to their perceptual correlates. The functions obtained appear to accord well with those reported by Stevens (1957) and by Stevens and Galanter (1957) for basic dimensions involving relatively simple stimulus patterns. The present linear size and sharpness plots may be described as power functions with exponents of 1.0. Contrast, described most adequately by a hyperbolic equation, may also be described roughly as a power function with an exponent of .63–.69. Stevens (1957), considering the power function to be a first approximation to the psychophysical law, listed empirically derived exponents for a number of basic dimensions, including brightness (.3–.5), visual length (1.1),

and visual area (.9–1.15). The correspondence of these values with those obtained in the present study for contrast and size, respectively, is stiking. The implications of this agreement for the generality and usefulness of the method are clear, as noted earlier.

In one respect the present data fail to bear out previous findings with regard to the method of magnitude estimation. Stevens (1956) has shown that the slope of the psychophysical function for loudness is relatively impervious to changes in the amplitude of the modulus (within limits). If scales based upon magnitude-estimation procedures are to be considered ratio scales, such consistency in number usage is demanded. Failing this, the various functions would possess different points of zero magnitude, a condition obviously contrary to the requirements of a ratio scale. Such a condition appears to characterize a large portion of the present data. In Exp. I significant slope changes were incurred for both the size and sharpness functions as the modulus value was shifted by a single step at the center of the stimulus range (see Fig. 2; Table 2). For contrast, however, such a shift in modulus had no noticeable effect on slope. Therefore, it must be concluded that the method of magnitude estimation does not yield ratio scales for perceived size or sharpness, but does yield such scales for contrast. For all three continua, however, the requirements for equal-interval scales are met. These findings are of particular interest because they emphasize the point that the method of magnitude estimation does not, in and of itself, guarantee ratio scaling of quantitative continua. This, of course, does not detract from the method. As Torgerson (1958) has pointed out, it is actually the scale and not the method which is being validated for ratio or interval status in a test such as that conducted above. In view of the fact that not all of the scales developed in the present study possess ratio properties, however, it would appear that an evaluation of other scales obtained using the magnitude estimation method would be profitable. If all such scales were to meet the requirements of ratio measurement, as might well be the case, the present findings could be attributed to the complexity of the stimuli.

Two of the investigations (Exp. II and III) were concerned with the independence of judgments of size, sharpness, and contrast. As expected, certain interactions did appear. Size judgments, for example, were increased by reducing sharpness and decreased by reducing contrast. Reducing both together, however, had no appreciable effect on size judgments. Neither, for that matter, did reducing both size and sharpness have an appreciable effect on judgments of contrast. Drastic reductions in size and contrast, however, did bring about a complete disruption of sharpness judgments. Whether less drastic reductions in size and contrast can be made without disrupting sharpness judgments remains to be determined. Similarly, the influence of individual size and sharpness changes on contrast judgments and of individual size and contrast changes on sharpness judgments is still unknown. Despite the fact, then, that only a sample of the possible interactions among size, sharpness, and contrast were explored in the present work, it is apparent that the perceptual dimensions involved are not completely dependent upon their specific physical counterparts. On the other hand, the obtained interaction effects were sufficiently small to justify the conclusion that quantifiable perceptual continua do exist which correspond closely to the three physical dimensions. The situation is probably analogous to that of the perceptual hue dimension which is dependent primarily upon the wavelength of light and, to a lesser extent, upon light intensity and saturation.

Summary

Three dimensions considered to be of importance in the recognition of complex visual stimuli were scaled psychophysically using the method of magnitude estimation. The dimensions investigated were size, blur,

and contrast, and the stimulus figures used were letters of the alphabet. Each dimension was scaled (*a*) with the other two at optimum-legibility values and (*b*) with the other two at recognition threshold values. In addition, size was scaled under minimum values of sharpness and contrast taken individually.

The results indicated that size and sharpness judgments are linearly related to their physical counterparts (letter height and blurredness), whereas the function for contrast is most adequately described by a hyperbolic equation. Reducing two of the three dimensions to legibility threshold does not appear to exert a serious influence upon judgments of size or contrast, although it does effectively disrupt sharpness judgments. Whether a change in the sharpness *function* also results from the reduction of size and contrast is difficult to ascertain from the present data owing to the variability of judgments obtained under these conditions.

Reduction to recognition threshold of contrast and sharpness taken individually has a small, but statistically reliable, effect on size judgments, such that reduced sharpness increases the apparent size of the stimulus figure and reduced contrast decreases it.

From a methodological standpoint, the present data provide strong evidence for the usefulness of the method of magnitude estimation in scaling perceptual continua. This conclusion is based on the fact that the method yielded reliable psychophysical functions for size, sharpness, and contrast under a wide variety of complex stimulus conditions. Although two of the obtained continua are found to lack the necessary characteristics of ratio scales, this fact does not detract from the validity of the magnitude estimation method. Rather, it suggests that all scales obtained using this technique should be tested to determine their measurement properties.

References

Berger, C. Stroke-width, form and horizontal spacing of numerals as determinants of the threshold recognition. *J. appl. Psychol.*, 1944, **28**, 208–231.

Crook, M. N., Harker, G. S., Hoffman, A. C., & Kennedy, J. L. Effects of amplitude of apparent vibration, brightness, and type size on numeral reading. *USAF AMC tech. Rep.*, 1950, No. 6246.

Fry, G. A., & Enoch, J. M. Human aspects of photographic interpretation. Third Interim Research Rep., Contract No. AF 30(602)-1580, RADC, Rome, New York, 1957.

Fry, G. A., & Enoch, J. M. The relation of blur and grain to the visibility of contrast borders and gratings. *USAF RADC tech. Note*, 1959, No. 59–237.

Howell, W. C., & Kraft, C. L. Size, blur, and contrast as variables affecting the legibility of alpha-numeric symbols on radar-type displays. *USAF WADC tech. Rep.*, 1959, No. 59–536.

Lansdell, H. Effects of form on the legibility of numbers. *Canad. J. Psychol.*, 1954, **8**, 77–79.

Schapiro, H. B. Factors affecting legibility of digits. *USAF WADC tech. Rep.*, 1952, No. 52–127.

Spragg, S. D. S., & Rock, M. L. Dial-reading performance as a function of brightness. *J. appl. Psychol.*, 1952, **36**, 128–137.

Stevens, S. S. The direct estimation of sensory magnitudes: Loudness. *Amer. J. Psychol.*, 1956, **69**, 1–25.

Stevens, S. S. On the psychophysical law. *Psychol. Rev.*, 1957, **64**, 153–181.

Stevens, S. S. Problems and methods in psychophysics. *Psychol. Bull.*, 1958, **55**, 177–196.

Stevens, S. S., & Galanter, E. H. Ratio scales and category scales for a dozen perceptual continua. *J. exp. Psychol.*, 1957, **54**, 377–411.

Tinker, M. A. Relative legibility of letters and digits. *J. gen. Psychol.*, 1928, **1**, 472–496.

Torgerson, W. S. *Theory and methods of scaling.* New York: Wiley, 1958.

(Received December 4, 1959)

Journal of Experimental Psychology
1961, Vol. 61, No. 1, 40-44

THE SUBJECT EFFECT IN JUDGMENTS OF SUBJECTIVE MAGNITUDE[1]

F. NOWELL JONES AND MAXINE J. MARCUS

University of California, Los Angeles

A century has passed since the publication of Fechner's *Elemente der Psychophysik*—a century marked by sporadic attacks, defenses, and suggestions for the modification of Fechner's Law relating stimulus magnitude to magnitude of sensation. The most recent attempt to state a general psychophysical law intended to replace Fechner's is due to Stevens. Simply put, Stevens asserts that "equal stimulus ratios produce equal sense ratios" (Stevens, 1957, p. 153).

This assertion is based mainly upon the results obtained by means of the method of magnitude estimation. In this method, *S* is presented with a series of stimuli differing in physical magnitude, say lights of different intensity, and is instructed to judge their magnitudes by assigning to each a numeral. Usually, a standard stimulus of some arbitrarily preassigned numerical value is also presented to *S*, and the numerals he assigns to the comparison stimuli are presumably in proper ratio to the standard and to each other for the representation of the subjective magnitudes corresponding to the physical stimuli involved. Typically, the results from 20 or 30 *S*s are pooled, and the subjective scale value assigned to each stimulus magnitude is obtained by averaging over *S*s. The most usual average is the median, but sometimes the geometric mean is used. The empirical function used to describe the relationship which results from this averaging is $R = aS^b$, where S is stimulus magnitude and R is the average numeral assigned, i.e., subjective magnitude. Although tests of goodness of fit have not been made, correlations between stimulus and response appear to be high (Jones, 1958a, 1958b).

Despite its general usefulness, averaging over *S*s may conceal important information. For example, if one accepts Gulliksen's test of the Stevens formulation he will need to know the individual results (Gulliksen, 1959). Or one may ask whether or not the empirical constants obtained by averaging over *S*s are the same as those that would be obtained by repeated measurement of the same *S*. If not, is averaging formally justified? And does a given *S* tend to use numerals in a consistent way which he carries from modality to modality? This experiment is designed to give at least preliminary answers to these questions.

METHOD

Three modalities were selected which were expected to yield rather different empirical constants. These were weight, taste, and smell. Saltiness was selected to represent taste, and the odor of benzene was chosen for smell. A series of seven stimuli was prepared for each modality. Within each modality, successive stimuli differed by a constant ratio. The extreme values were: for weight, 94 and 940 gm.; for taste, .03 and 1.02M; for smell, .05 and 87 mm. Hg.

Scaling was carried out using the usual procedures of magnitude estimation. Specifically, the middle stimulus of each modality was designated as the standard, and was arbitrarily assigned the value "5." The *S* was required to judge, by the assignment of a numeral he deemed appropriate, all seven stimuli of a given series against the appropriate standard which was presented before each comparison. In a 1-hr. session *S* gave two judgments of every stimulus. The order of presentation of modalities was balanced over *S*s, and the second run through a series was the reverse of the first. The order of stimuli within the modalities was in no case

[1] The work reported here was made possible by grants from the United States Public Health Service and the University of California, Los Angeles, Committee on Research. We wish to thank N. H. Anderson for many helpful comments.

continuously increasing or decreasing and was the same for all *Ss*.

The weights consisted of unmarked cardboard tubes, differentially weighted. The salt solutions were presented in about 6 ml. doses in shot glasses. A distilled water rinse followed each sample. The odors were prepared by dilution in mineral oil, and were presented in wide-mouthed bottles for sniffing. Central air conditioning kept the experimental room free of smog and within a degree of 74° F.

The *Ss* were drawn largely from elementary classes in psychology in which service was required. Fifty-two *Ss* completed the experiment. Three *Ss* were dropped since they gave one or more zero magnitude judgments which made the fitting of curves based on a logarithmic transformation impossible.

Results

It was decided to concentrate on the linear component of the log-log relationship for two reasons: (*a*) the theoretical function under consideration includes only this component, and (*b*) most of the variance will be accounted for even if some curvilinearity is actually present. Individual curves of the form $\log R = \log a + b \log S$ were fitted to each of the two orders for each modality for each *S* for a total of 294 curves. It should be noted that the equation stated above is the logarithmic form of $R = aS^b$. The computations were made by the IBM 709 computer at the Western Data Processing Center, UCLA. Correlation of the individual results with the physical magnitudes of the stimuli was reasonably good; the median coefficient was .95. It is not possible to be certain, despite these high correlations, that there was not significant curvilinearity in the data.

TABLE 1

Median and Extreme Exponents (b) in $R = aS^b$

Measure	Weight	Taste	Smell
Median	1.7	1.05	.56
Q_3	2.09	1.27	.93
Q_1	1.32	.54	.47

TABLE 2

Analysis of Variance of Exponents for Weight, Taste, and Smell

Source	*df*	*MS*	*F*
Order (O)	1	.01	.06
Modality (M)	2	16.18	9.7**
Subjects (*S*)	48	1.79	7.7**
O × M	2	.22	1.29
O × *S*	48	.23	1.35
M × *S*	96	1.71	10.05**
O × M × *S*	96	.17	

** $P < .001$.

Table 1 gives the median and upper and lower quartile exponents—the b in the equations of the preceding paragraph—for each modality. The basis of our interest in this constant may be appreciated when it is considered that in the power function $R = aS^b$ the coefficient a represents merely a unit of measurement. An analysis of variance was performed upon the individual exponents and is summarized in Table 2. It can be seen that the main effects of Modality and *Ss* are highly significant while that of Order is not. The modality effect was to be expected on the basis of past work showing the dependence of exponent on modality (cf. Stevens, 1957). The *S* effect is very interesting, however, since it implies a consistency in the use of numerals by a given *S* from modality to modality.

The only significant interaction is between Modality and *S*. The interpretation of this interaction requires further analysis. If the obtained exponents represent a multiplicative effect of *S* and Modality, the usual analysis of variance, depending as it does on an additive model, is inappropriate. This point is discussed by Williams (1952) who gives an analysis

of variance assuming a multiplicative relationship. Because of the rather large N, a somewhat simpler use of the multiplicative model seemed permissable. Frequency distributions of the exponents for each modality were prepared. The distributions for taste and weight were nearly normal, while that for smell was, for unknown reasons, markedly skewed. Since it is not clear what effect skewness might have upon the multiplicative model, taste and weight were selected for further analysis. In order to determine if the omission of smell would alter the outcome of the original analysis of variance, a similar analysis of the exponents for taste and weight only was performed. The same pattern of significant effects was formed, it being especially noteworthy that the Modality by S interaction was still significant at the .005 level. The multiplicative model was next applied by multiplying each taste exponent by a constant obtained by dividing the mean taste exponent by the mean weight exponent. Analysis of the resulting adjusted exponents, given in Table 3, shows that only the S effect is significant. The modality effect has disappeared because of the adjustment. The most interesting result, however, is the disappearance of the Modality by S interaction, since this is not an automatic consequence of the method of adjustment. This disappearance implies that a given exponent can be represented as the product of two constants: one constant for the given modality, and a second constant for the individual S.

TABLE 3

ANALYSIS OF VARIANCE OF EXPONENTS FOR WEIGHT AND TASTE, ADJUSTED TO A COMMON AVERAGE

Source	*df*	*MS*	*F*
Order (O)	1	.027	
Modality (M)	1	0	
Subjects (S)	48	.894	2.75*
O × S	48	.325	1.25
M × S	48	.395	1.52
O × M	1	.845	3.26
O × M × S	48	.259	

* $P < .005$.

DISCUSSION

Under the conditions of this experiment, individual Ss judged sensory magnitudes consistently from modality to modality. This is tantamount to saying that some Ss produce a relatively wide range of numerals, others a relatively narrow range. The outcome is that those who use a wide range produce data fitted by a curve with a relatively large exponent, and when this tendency is carried from modality to modality it leads to a positive correlation of exponents from the different modalities. Since the multiplicative model was verified by the adjusted analysis of variance, the power law must be modified to include a constant characteristic of the individual S. Thus, if c_i is appropriate to S_i, we may write $R = aS^{bc_i}$, for a given modality. The form of the relationship is not, however, altered by this modification.

The introduction of what may be termed a "personal exponent" has interesting implications concerning the nature of the subjective scales obtained from magnitude estimation. For example, Gulliksen (1959) has proposed a matrix test of the power law. Individual judgments are to be entered in a S (rows) by Stimulus Magnitude (column) matrix. If the rank of this matrix is one, the power law is verified. To say that the rank of the matrix is one means that it is reducible to a column matrix of individual S coefficients postmultiplied by a row matrix of stimulus magnitudes to a power determined by the characteristic exponent for the given modality. If this suggestion is correct, then the exponent should be the same for all Ss for a given modality, since the

coefficients represent merely a change of unit. It is evident that our data do not fulfill this condition. It is possible, of course, that the data for each *S* follow the power law. Even if this is true, however, the meaning of a general power law based on data averaged over *S*s must be given closer study.

Several authors have considered the effect of averaging upon the interpretation of data, and have discussed the forms of functional relationship which permit on formally acceptable grounds the taking of arithmetic averages (Bakan, 1954; Estes, 1956; Sidman, 1952). A function of the form $R = aS^b$ does not permit of averaging. Specifically, this function will not meet the tests suggested by Estes since the exponents do not enter into an average in a linear manner. But if we write the power function in logarithmic form a way out suggests itself. Log $R = \log a + bc_i \log S$ is a linear equation and therefore averageable. The average exponent will be equal to $(bc_i + bc_j + \cdots + bc_N)/N$. To accomplish the suggested transformation one may either take logarithms of the individual data or take the geometric mean of the data at each abscissa point. If one is content with taking the median logarithm across *S*s taking the median raw score introduces some but not considerable distortion. This follows because in the process of curve-fitting it is necessary to find the logarithm of the average score anyway, and the median of a set of scores logarithmically transformed will not be very different from the logarithm of the median raw score.

If the proper average is used—and even medians will not be very misleading—the group exponent will be representative of a modality exponent multiplied by an average personal exponent without distortion of the underlying function. Admitting the possibility of individual differences weakens the force of Gulliksen's test, since all that is required is that each *S* have an empirical function of proper form but not necessarily the same exponent. It will be necessary to take more extensive measures upon each *S* and analyse the data for curvilinearity before the power law can be tested in its general form. This we have not attempted to do in this experiment.

It is not clear to what may be ascribed the obtained differences in the usage of numerals. Either *S*s with relatively large personal exponents do indeed have more intense subjective magnitudes or they have different conceptions of the number system. Since subjective magnitude cannot, of course, be measured directly, this issue cannot be resolved without further experimentation to determine to what extent individual differences may be altered by the manipulation of *S*s' experience with numbers. To put it slightly differently, one must try to determine whether the obtained differences are to be ascribed to what may be called perceptual differences or to differences in habits of response. The elucidation of this problem must await further investigation.

Summary

Forty-nine *S*s judged the subjective intensities of a series of weights, tastes, and smells by the method of magnitude estimation. Individual curves of the form $R = aS^b$ were were fitted for each *S* for each modality and for each of two runs, here referred to as orders. Analysis of variance of the exponents, b, so obtained reveals: (*a*) significant primary effects of *S*s and Modalities, (*b*) a significant interaction between Modality and *S*s, (*c*) the removal of the interaction in the case of taste and weight by adjusting the average b for the two modalities to equality. A multiplicative model being appropriate, the power law of Stevens is revised to include an individual exponent, hence $R = aS^{bc_i}$, where c_i is appropriate to a given *S*. A discussion of averaging to obtain group curves reveals that either a logarithmic transform prior to averaging or the geometric mean are logically correct, although the use of medians should not introduce great distortion.

References

Bakan, D. A generalization of Sidman's results on group and individual functions, and a criterion. *Psychol. Bull.*, 1954, **51**, 63–64.

ESTES, W. K. The problem of inference from curves based on group data. *Psychol. Bull.*, 1956, **53,** 134–140.

GULLIKSEN, H. Mathematical solutions for psychological problems. *Amer. Scientist,* 1959, **47,** 178–201.

JONES, F. N. Scales of subjective intensity for odors of diverse chemical nature. *Amer. J. Psychol.*, 1958, **71,** 305–310.

JONES, F. N. Subjective scales of intensity for three odors. *Amer. J. Psychol.*, 1958, **71,** 423–425.

SIDMAN, M. A note on functional relations obtained from group data. *Psychol. Bull.*, 1952, **49,** 263–269.

STEVENS, S. S. On the psychophysical law. *Psychol. Rev.*, 1957, **64,** 153–181.

WILLIAMS, E. J. The interpretation of interactions in factorial experiments. *Biometrika*, 1952, **39,** 65–81.

(Received December 8, 1959)

Journal of Experimental Psychology
1961, Vol. 61, No. 1, 45–50

ANALYSIS OF THE ROLE OF OVERLEARNING IN DISCRIMINATION REVERSAL [1]

M. R. D'AMATO AND HARRY JAGODA

New York University

Several studies using rats in a brightness discrimination situation have shown that postcriterion learninging (overlearning) facilitates discrimination reversal learning (Capaldi & Stevenson, 1957; Pubols, 1956; Reid, 1953). Relatively complex processes have been advanced to account for this result, such as that the rat learns a "response of discriminating" (Reid, 1953, p. 107), or that "a change in a problem is more readily discriminated and extinction is faster following overtraining" (Capaldi & Stevenson, 1957, p. 197), presumably leading to faster reversal. In a previous paper (D'Amato & Jagoda, 1960) the present authors offered a simpler interpretation, an experimental test of which forms the present report.

In essence, our view is that an essential component of simple discrimination learning is the development of avoidance tendencies toward S_-; the establishment of approach tendencies toward S_+ is, in our opinion, ordinarily a process of secondary importance. Reversal learning, on the other hand, critically requires the extinction of avoidance tendencies developed toward the former S_-, and it is this factor that makes reversal learning so difficult (D'Amato & Jagoda, 1960; Warren & Kimball, 1959). Relating this hypothesis to the overlearning results, we have maintained that during postcriterion training, experience with S_- is virtually terminated (i.e., *Ss* in general make very few errors), resulting in a reduction of avoidance tendencies toward the negative stimulus. This reduction is possibly brought about in part through generalization of approach tendencies from S_+. In any event, our prediction was that if *Ss* were compelled during overlearning to have a reasonable number of experiences with S_-, the facilitative effect of postcriterion training would disappear.

As a test of this hypothesis, three groups of rats were trained on a brightness discrimination problem: Group C was reversed immediately upon reaching criterion, Group C-200 was given 200 overlearning trials, 20% of which were forced to S_+, and Group E had 20% of its overlearning trials forced to S_-. Because of the continued experience with S_- during the postcriterion trials, Group E was predicted to take significantly longer than Group C-200 to learn the reversal. In line with the previous overlearning studies, Group C-200 was expected to learn the reversal faster than Group C.

METHOD

Subjects.—The *Ss* were 51 experimentally naive albino rats, 70–75 days of age at the start of the study. Four *Ss* were discarded because of programing errors or failures to run; two *Ss* were discarded in order to equate groups with respect to size and acquisition means.

Apparatus.—Four units of a previously described (D'Amato & Jagoda, 1960) automatic Y maze discrimination apparatus

[1] This research was supported by Research Grant M-2051 from the National Institute of Mental Health, National Institutes of Health, United States Public Health Service.

(ALDA) were used. Briefly, ALDA consists of a symmetrical Y maze with remote programing and recording equipment. The use of a symmetrical Y maze permits the same arm to be used as a "goal" and "starting" arm, allowing for a sequence of automatically controlled trials. For example, in a brightness discrimination problem the (opaque) doors open and *S* chooses between a bright and a dimly illuminated arm. The arm of his choice becomes, on the next trial, the "starting" arm, the discriminanda appearing in the remaining two arms. Each trial runs off in this manner in accordance with a pre-established program until the day's trials are completed.

Pretraining.—Four days prior to the beginning of discrimination training, all *S*s were put on a standard 23-hr. water deprivation regimen, food being available ad lib. On the last day of deprivation training each *S* was permitted to drink 5 dipperfuls of water in a darkened arm of the maze; 3 dipperfuls were allowed on the first day of discrimination training.

Discrimination training.—All *S*s were run 20 trials per day on a brightness discrimination problem until the criterion of 18 correct choices out of any 20 successive free trials was met. From Day 2 on, 20% of the daily 20 trials were forced by locking closed the door to one of the arms. One trial in each block of five was forced, the forced trials being randomly chosen with the restriction that half of the daily forced trials were to S_+ and half to S_-. Five such forcing programs were developed and used during discrimination training and overlearning.

For all *S*s, S_+ was the brightly lighted arm (10.0 ft-c, with the photometer target resting on the floor of an arm, its leading edge 2.5 in. from the closed door); S_- was the dimly lighted arm (no reading was possible on the meter used, though a discernible low level of illumination was present). The starting stimulus was a flashing light of moderate intensity.

On a correct choice the sequence of events was as follows. The *S* was permitted 1.25 sec. ± .25 sec. drinking (reward interval), timed from *S*'s first lap of water, after which the dipper retracted out of sight. The discriminandum, S_+ in this case, remained on for an additional 8 sec. ("stimulus" interval). Then a 57 sec. ± 3 sec. (intertrial) interval followed, during which all lights in the maze were extinguished. Finally, a flashing light during a 4.5 sec. ± .5 sec. ("starting") interval signaled the beginning of the next trial. On an incorrect choice the sequence of events was identical, with the dipper retracted during the initial 1–1.5 sec. interval.

During discrimination training (from Day 2 on) and during overlearning and reversal learning, the estimated average duration of water deprivation at the start of a day's trials was 22 hr.; range: 21–23 hr. The temperature range during the experiment was 72–78° F; relative humidity 50–65%.

Overlearning and reversal.—Upon reaching criterion *S*s were randomly assigned to three groups of 15 *S*s each, equated with respect to mean numbers of trials to criterion and variances. The *S*s of the control group, Group C, were started on reversal training the day after reaching criterion. An extended period of overtraining first intervened for Groups C-200 and E. Group C-200 was given 200 overlearning trials, 20 per day, the important feature of these trials being that 20% of the daily trials were forced to S_+ (none was forced to S_-). Seven *S*s of Group E (Subgroup E-200) were also given 200 post-criterion trials, with the critical difference that the forced daily trials (20%) were all to S_-; the remaining 8 *S*s of this group (Subgroup E-240) received 12 days of overlearning, i.e., 240 trials in all, with 20% of the daily trials also forced to S_-.

Reversal training was identical in procedure to discrimination training, with the exceptions that all 20 daily trials were free and the brightly lighted arm, the former S_+, was now S_-. All *S*s were run until the criterion of 18 correct responses out of 20 successive trials was met.

Results

Discrimination.—The mean numbers of free trials to criterion for Groups C, C-200, and E were, in order, 49.60, 51.07, and 50.73; the respective *SD*s were 12.00, 15.19, and 13.33. The three groups turned out also to be well equated with respect to mean numbers of errors to criterion (24.47, 24.20, 25.53, in the same order).

Overlearning.—Some indication of the efficaciousness of the forced trials to S_- may be derived from a consideration of the number of errors made by *S*s of Groups C-200 and E during overlearning. If a forced trial to S_- is at least partially as

TABLE 1

MEAN NUMBER OF EXPERIENCES WITH S_+ AND S_- DURING OVERLEARNING

Group	S_+	S_-	N
C-200	190.9	5.1	15
Subgroup E-200	157.1	42.6	7
Subgroup E-240	188.1	51.8	8

effective as a free choice of S_-, then *Ss* of Group E should make fewer errors during overlearning than *Ss* of Group C-200, in which all forced trials were to S_+.

The mean number of total errors for the first eight days of overlearning (128 free trials)[2] was 4.67 for Group C-200 and 2.67 for Group E ($t = 1.71$; $.05 < P < .10$). That any errors at all occur in Group E constitutes evidence that the 32 forced trials to S_- were not equivalent to free choices of that stimulus. Our claim that experience with S_- is slight during postcriterion training is substantiated by the fact that *Ss* of Group C-200 average fewer than 4% errors during the first eight days of overlearning.

Table 1 presents the mean numbers of experiences with S_- and S_+, free and forced trials included. Note that Group C-200 and Subgroup E-200, groups in which the number of postcriterion trials was controlled, differ widely with respect to the number of experiences with both S_- and S_+. Thus, if in accordance with our hypothesis Group C-200 reversed faster than Subgroup E-200, this superiority could not unambiguously be assigned to Group C-200's fewer experiences with S_-. As a control for the number of experiences with S_+, Subgroup E-240 was given 40 additional overlearning trials, and as a result this subgroup was roughly equated with Group C-200 on this variable.

Reversal.—Considering first the results of Subgroups E-200 and E-240, the mean numbers of trials to reversal criterion were, in order, 118.43 and 130.50; the mean numbers of errors to criterion were, in the same order, 75.71 and 88.50. Since t tests on both trials to criterion ($t = .90$) and error scores ($t = 1.14$) failed to reveal significant differences between the subgroups, only the combined data (Group E) will be considered in further analysis.

The group means for both trials to criterion and errors to criterion are presented in Table 2. Analysis of variance of the trials to criterion data, in a groups by levels design, produced an F for groups of 3.20 ($df = 2, 36$; $P \approx .055$). The major comparison of interest, i.e., the difference between the means of Groups C-200 and E, is clearly significant: $t = 2.45$, $P = .02$. In verification of previous studies, Group C-200 reversed faster than Group C ($t = 1.76$; $.05 < P < .10$). The difference between the means of Group C and E did not prove significant.

TABLE 2

GROUP MEANS TO REVERSAL CRITERION

Group	Trials to Criterion	Errors to Criterion
C-200	92.00	58.53
C	115.60	70.07
E	124.87	84.14

[2] In error one *S* from Group C-200 was placed on reversal after only 160 overlearning trials and another after 180 trials. Since ease of reversal is positively related to the amount of overlearning, these errors should tend to increase the reversal average of this group, thus working against our hypothesis. Due to this inadvertence and an occasional failure in print-out, the mean experiences with S_- and S_+ fail to sum to the proper totals in Table 1.

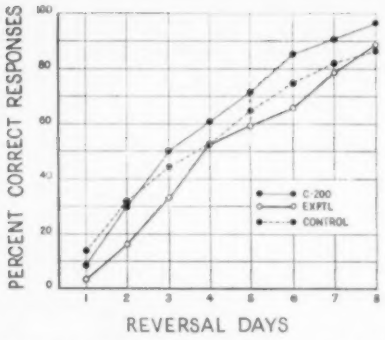

FIG. 1. Percentage of correct responses during the first eight days of reversal training.

A similar analysis applied to the error scores presented the same general picture. With the error measure the F for groups alone was significant ($F = 3.94$, $P < .05$). The Group C-200 vs. Group E comparison was clearly significant ($t = 3.24$, $P < .005$). The difference between the means of Groups E and C yielded a t of 1.78 ($.05 < P < .10$), while the Groups C vs. Group C-200 comparison fell considerably short of significance ($t = 1.46$).

Figure 1 presents the graphed data for the first eight days of reversal training, by which time all *S*s of Group C-200 had achieved criterion. (For purposes of the graph, *S*s that reached criterion before the eighth day were assigned a score of 100% on succeeding days.)

Finally, as previously reported (D'Amato & Jagoda, 1960), discrimination and reversal learning were significantly related in Group C (rho = .696 for the trials to criterion measure, and .539 for the errors to criterion score). As might be expected, the overlearning experience served greatly to attenuate this relationship in Groups C-200 and E.

DISCUSSION

Our results clearly show that the facilitative effect of overlearning on discrimination reversal learning can parsimoniously be attributed to the virtual termination of experience with S_- during the postcriterion trials, and recourse to more complex interpretations appears unnecessary.[3] It will be recalled from Table 1 that differential experience with S_+ during the postcriterion trials was ruled out as a factor with respect to Group C-200 and Subgroup E-240. Since the subgroups of Group E did not differ significantly between themselves, we must conclude that the differential experience with S_- is responsible for the significance obtained between the reversal means of Group C-200 and Group E.

We have hypothesized that as a consequence of its minimal experience with S_- during overlearning, Group C-200 suffers a reduction in its avoidance tendencies toward this stimulus, resulting in a facilitation of reversal learning. We recognize two important means by which avoidance of S_- could be reduced in this situation; one is direct, while the other is rather indirect. The latter is the generalization, during overlearning, of approach tendencies from S_+ to S_-, which might serve to reduce avoidance of S_- to a point where an occasional error would maintain avoidance of S_- at the minimum intensity necessary for adequate performance of the discrimination. In Group E, however, the forced trials to S_- guarantee a level of avoidance of S_- considerably above this minimum value.

More directly, on the other hand, the infrequent experience with S_- could, per se, contribute to a reduction in the

[3] In a recent report (Birch, Ison, & Sperling, 1960) an even simpler explanation of the overlearning effect has been suggested, namely, that extinction of the approach response to the positive stimulus is a nonmonotonic function of the number of acquisition trials. Hence, overtraining facilitates reversal learning because it leads to faster extinction of the formerly correct response.

avoidance of that stimulus, as a kind of forgetting. If that were the case, mere delay between acquisition of the discrimination and reversal training should facilitate reversal learning. It is interesting that in none of the relevant studies cited above was this factor evaluated. On the other hand, there exist at least three studies that report facilitation of reversal learning as a function of increased delay between acquisition and reversal training (Bunch, 1939; Murofushi, 1957; Stevenson & Weir, 1959). The relative contribution of the direct and indirect factors remains to be evaluated.

Although the present study and its implications are most relevant for visual discrimination situations, we feel that our analysis is valid even for position habits (e.g., Pubols, 1956). Thus the question arises whether our interpretation may not be appropriate for the results recently reported by Bruner, Mandler, O'Dowd, and Wallach (1958). These investigators trained rats on a "four-unit straight alley T maze" on an RLRL (or LRLR) pattern, reversing *Ss* after various degrees of overlearning (their drive variable is not relevant here). Except for *Ss* trained and reversed under high drive, they found the usual overlearning effect, i.e., overlearning facilitated reversal. They rather intrepidly assigned this facilitation to the rats' "learning a principle" (of single alternation) on the basis of "generic recoding," terms that imply processes of a rather high order of complexity. It is our hunch that the very same simple process is operating here as in our study: owing to the few errors that are made in overlearning, there results a reduction in the avoidance of the closed door at each of the four choice-points, thus rendering reversal learning easier.

If, as seems likely from the mounting evidence (e.g., Lachman, 1959; Warren & Kimball, 1959), lower mammals (for supporting evidence at the primate level see Harlow and Hicks [1957]), particularly naive *Ss*, learn what not to approach, i.e., learn to avoid S_-, how shall one interpret this behavior? One possibility is that, given the usual motive-incentive conditions operating in experimental situations, approach responses to both discriminanda are relatively high in the response hierarchy. Under an appetitive drive and appropriate incentive, animals quite "normally" approach stimuli not otherwise aversive. Hungry cats quite naturally reach *toward* wells where there might be food; thirsty rats quite naturally walk *into* alleys where they might find water, and so on. From this point of view, approach to S_+ hardly needs to be conditioned since it is already present from the start, almost as an unconditioned response. The major requirement of the animal, it appears, is to learn to check his approach responses to the wrong stimulus, i.e., to learn to avoid S_-.

Summary

Three groups of rats were reversed on a simple brightness discrimination after various conditions of overlearning. Group C-200 received 200 overlearning trials, 20% of which were forced to S_+. Group E received a similar number of postcriterion trials, with the critical difference that 20% of these trials were forced to S_-. The control group, Group C, was reversed after zero overlearning trials. Group C-200 learned the reversal faster than either Group C or Group E, the latter group being the slowest of the three. The relevance of these and other results for an "avoidance" interpretation of simple discrimination learning was discussed.

References

Birch, D., Ison, J. R., & Sperling, S. E. Reversal learning under single stimulus presentation. *J. exp. Psychol.*, 1960, **60**, 36–40.

Bruner, J. S., Mandler, J. M., O'Dowd, D., & Wallach, M. A. The role of overlearning and drive level in reversal learning. *J. comp. physiol. Psychol.*, 1958, **51**, 607–613.

Bunch, M. E. Transfer of training in the mastery of an antagonistic habit after varying intervals of time. *J. comp. Psychol.*, 1939, **28**, 189–200.

Capaldi, E. J., & Stevenson, H. W. Response reversal following different amounts

of training. *J. comp. physiol. Psychol.*, 1957, **50,** 195–198.

D'AMATO, M. R., & JAGODA, H. Effects of extinction trials on discrimination reversal. *J. exp. Psychol.*, 1960, **59,** 254–260.

HARLOW, H. F., & HICKS, L. H. Discrimination learning theory: Uniprocess vs. duoprocess. *Psychol. Rev.*, 1957, **64,** 104–109.

LACHMAN, R. Drive level and forced choice response schedules in brightness discrimination. Unpublished doctoral dissertation, New York Univer., 1959.

MUROFUSHI, K. The effect of cue on reversal learning after varying periods of rest. *Psychologia*, 1957, **1,** 37–46.

PUBOLS, B. H., JR. The facilitation of visual and spatial discrimination reversal by overlearning. *J. comp. physiol. Psychol.*, 1956, **49,** 243–248.

REID, L. S. The development of noncontinuity behavior through continuity learning. *J. exp. Psychol.*, 1953, **46,** 107–112.

STEVENSON, H. W., & WEIR, M. W. Response shift as a function of overtraining and delay. *J. comp. physiol. Psychol.*, 1959, **52,** 327–329.

WARREN, J. M., & KIMBALL, H. Transfer relations in discrimination learning by cats. *J. comp. physiol. Psychol.*, 1959, **52,** 336–338.

(Received December 14, 1959)

Journal of Experimental Psychology
1961, Vol. 61, No. 1, 51–56

FIGURAL AFTEREFFECTS IN KINESTHESIS: EFFECTS OF OBJECT WIDTH AND REPEATED PRESENTATIONS[1]

ERIC G. HEINEMANN

Vassar College

If a person holds a block of wood (the inspection or I-object) between the thumb and fingers for a few minutes and subsequently attempts to estimate the widths of other blocks (test or T-objects), he will in general underestimate the width of T-objects that are narrower than the I-object, and overestimate the width of T-objects that are wider than the I-object. This is the figural aftereffect, first described in this form by Köhler and Dinnerstein (1947).

The experiments to be reported are concerned with the manner in which the size of the figural aftereffect varies with the relative width of the I- and T-objects, and the possible growth of the aftereffects over periods of several days.

When these experiments were begun no quantitative information was available on the effects of the width of the I- and T-objects. Since then, Charles and Duncan (1959) have published data showing how changes in the width of the T-object affect the size of the aftereffects for T-objects that are narrower than the I-object. In the present experiments measurements were made, under conditions fairly similar to Charles and Duncan's, both for T-objects that are narrower than the I-object and for T-objects that are wider than the I-object.

The second problem with which the present experiments deal grows out of the observation by Köhler and Dinnerstein (1947) that, if *S* is given only a brief exposure to the I-object on each of a number of successive days, the figural aftereffects keep growing larger for a period of from several days to several weeks; even though with massed exposure the effect apparently reaches a maximum after only a few minutes of exposure to the I-object. The aftereffects established with widely spaced inspection periods may persist for many months, according to Köhler and Dinnerstein (1947). These observations have been confirmed recently by Wertheimer and Leventhal (1958).

It is not really clear, however, that the effects dealt with in these investigations are actually figural aftereffects. To show why an unambiguous interpretation is not possible it is necessary to examine briefly the procedure used in these experiments. When measurements of the "apparent width" of the T-block are made, *S* feels the T-block with one hand, while simultaneously feeling a scale with the other. In both of the experiments cited the hand to be used on the T-block was "satiated" with a block that was narrower than the T-block, and the hand to be used on the scale was "satiated" with a block that was wider than the T-block. (After this satiation *S* selects as equal in width to the T-block a part of the scale that is wider than that so selected before satiation.) Köhler and Dinnerstein themselves pointed out that since both hands are being affected by satiation, this procedure cannot yield a straightforward measure of figural aftereffect. They did assume, however, that the aftereffects

[1] This research was supported by a grant from the National Science Foundation.

obtained with this method represent the sum of two effects—the increase in the apparent width of the T-block and the decrease in the apparent width of the scale.

If this assumption is correct, then the progressive shift in *S*'s matches to wider parts of the scale that results from repeated satiation does indeed reflect a figural aftereffect. But the available evidence does not exclude several other possible interpretations. For example, this shift would also occur if the "apparent width" of both the T-block and the scale gradually decreased (or increased), but at different rates.

In some unpublished experiments the present author was unable to find the cumulative growth of the figural aftereffect under conditions which differed from those described by the previous investigators cited only in that (*a*) only the hand to be used on the T-block was satiated, and (*b*) the T-block was narrower than the I-block. On the other hand, with T-blocks wider than the I-block, a cumulative growth of the tendency to overestimate the width of the T-block *was* found. In the hope of shedding some further light on this matter, the possible cumulative growth of the figural aftereffects was studied then for a large series of combinations of I- and T-block widths.

Method

Apparatus.—The apparatus was similar to that used by Köhler and Dinnerstein (1947). The essential elements are 11 test blocks and a scale. The T-blocks were rectangular blocks of smoothly finished wood, 8 ft. long and varying in width from 1.25 in. to 3.75 in., in steps of 0.25 in. The scale was a single block of wood, 8 ft. long, the width of which changed by discrete steps of $\frac{1}{16}$ in. from $\frac{1}{2}$ in. at one end of the block to $4\frac{1}{2}$ in. at the other end. Each "step" was marked with a code number which was used by *E* in recording the responses. The $2\frac{1}{2}$-in. T-block served also as the I-block in all experiments.

The T-blocks were placed on one table, the scale on another. These tables were placed parallel to each other and about 22 in. apart.

Subjects.—The *S*s were 12 women undergraduate students. Six of them were randomly assigned to the Experimental Group, the other six to the Control Group.

Procedure.—The *S*s were blindfolded throughout the experiment. All of them stood between the two tables in such a way that the scale was always on their right, the T- and I-blocks on their left.

The *S*'s primary task during the experiment was to select a step on the scale that appeared equal in width to a T-block felt with the other hand. They were instructed (and reminded when necessary) to hold the blocks and the scale between the thumb and the first two fingers, and to indicate a match only after having "bracketed" the point of apparent equality. While *S*s were making their judgments, they typically took several steps forward and backward between the two tables in order to explore a large part of the 8-ft.-long scale. After each judgment *S* removed her hands from the scale and T-block, whereupon *E* moved the scale backward or forward, in a haphazard fashion.

During the "satiation periods" *S* gripped the I-block between the thumb and first two fingers of the left hand and moved this hand slowly back and forth along the I-block.

For Group E the I-block and the T-block were placed parallel to each other on the left-hand table, about 4 in. apart with the I-block nearer to *S*. The *S*s were able to switch their grip from the I-block to the T-block within less than 1 sec.

The further details of the procedure for Group E are as follows: Each *S* served in 13 experimental sessions, one on each of 13 different days. During the first two days the "presatiation" points of subjective equality (PSEs) were determined. On each of these two days *S* matched each of the 11 test-blocks to the scale five times in succession. The 11 test-blocks were presented in a different random order to each *S* and on each day. There followed a period of two days during which no experimental sessions were held. Then the establishment and measurement of the figural aftereffects was begun. The *S* began each session with three 1-min. periods of satiation. These satiation periods were separated by 1-min. rest periods. Immediately following the end of the third 1-min. satiation period *S* made her first match.

From then on, each match *S* made was preceded by 20 sec. of satiation. The *S* made two successive matches of each T-block. About 15 sec. elapsed between the completion of the first match and the beginning of the second satiation period for a given T-block. After each set of two matches *S* was given a rest period of about 45 sec. while *E* changed T-blocks. The 11 T-blocks were presented in a different random order on each day.

The 11 satiation-test sessions were held on successive days except for a two-day interval of no experimentation between the Days 5 and 6.

The procedure for Group C was identical with that described except that all satiation periods were replaced by rest periods of equal length.

Results

For each *S*, 12 means were computed for each T-block. One of these, the presatiation mean, is based on the 10 matches made during the first two days. The other 11 are the post-satiation means (one for each day) and are based on 2 matches each. These are the basic data upon which Fig. 1 and 2 are based.

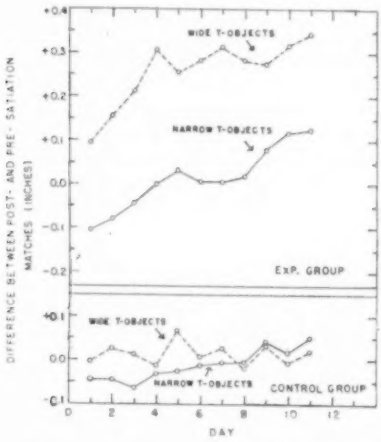

FIG. 1. Difference between post- and presatiation matches as a function of experimental day. Each point represents a mean value based on six *S*s and five test objects. See text for further details.

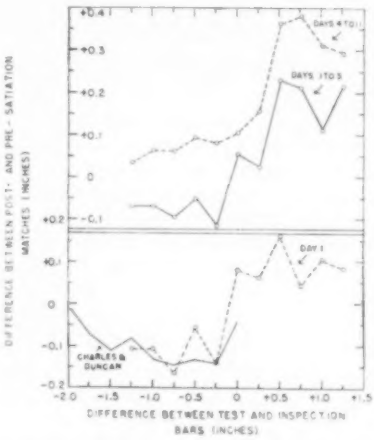
DAYS 4 TO 11
DAYS 1 TO 3
DAY 1
CHARLES & DUNCAN
DIFFERENCE BETWEEN POST- AND PRE-SATIATION MATCHES (INCHES)
DIFFERENCE BETWEEN TEST AND INSPECTION BARS (INCHES)

FIG. 2. Difference between post- and presatiation matches as a function of the difference in the widths of the test and inspection objects. Data from six *S*s.

Figure 1 shows the mean change in apparent width, as measured by the difference between post- and presatiation PSEs, as a function of the day on which the postsatiation PSE was determined. Each point represents a mean based on six *S*s and five T-blocks. For the function labeled "wide T-objects" the five T-blocks on which each mean is based are those wider than the I-block, for the function labeled "narrow T-objects" they are those narrower than the I-block.

The upper part of Fig. 1 shows that, for Group E, the apparent width of T-blocks both narrower and wider than the I-block increased over the 11-day period, and by approximately the same amounts. To evaluate the significance of this trend, the *t* test was used to test the hypothesis that the true difference between the means of the first five points and the last five points on each of the functions is zero. This hypothesis can be

rejected at the .05 level for both of the functions.

The trends shown are typical of the results of the individual *S*s. Though there are, of course, individual differences in the amount of change that occurred over the experimental period, the direction of change was the same for all *S*s.

The results for the control group were treated in the same way and are shown in the lower part of Fig. 1. There is no clear trend for the wide T-objects, but the curve for the narrow T-objects shows a slight rise. Application of the t test as described above for Group E shows that the difference between the means of the first five and the last five points is not significant ($P > .10$) for the wide T-objects, but is significant at the .05 level for the narrow T-objects.

Figure 2 shows how the aftereffect varies with the width of the T-block. In order to make possible a comparison with the results obtained by Charles and Duncan (1959) who used an I-block 2 in. wide, the difference in the width of the T- and I-block is laid off along the abscissa of this figure.

In the lower part of Fig. 2 are shown the results for the first day (means of 6 *S*s—dashed line) and also the results of Charles and Duncan (solid line). The results of the first day only were chosen for this comparison because they are based on satiation times most nearly comparable to those used by Charles and Duncan. The agreement between the two sets of data is quite good. The shape of this function bears out the generalizations that have been offered by Köhler and his collaborators. There is no figural aftereffect when the I- and T-objects are of approximately equal width. Objects narrower than the I-object are underestimated, objects wider than the I-object overestimated. The degree of over- or underestimation increases rapidly as the difference in the width of the I- and T-objects increases from zero, reaches a maximum near a difference of ½ in. and appears to decrease again for very large differences. This last feature is not very clear in these results, particularly not for the narrow T-objects, but the experiment of Charles and Duncan leaves no doubt of the fact that it occurs.

The upper portion of Fig. 2 shows the mean aftereffects for the first three days (solid lines) and the last eight days (dashed lines). This particular way of grouping the data was chosen because from the fourth day on measurements with the narrow T-objects on the average no longer showed any figural aftereffect, if this is defined as an underestimation of the narrow T-block. This can be seen in Fig. 1.

TABLE 1

AMOUNT OF AFTEREFFECT AS A FUNCTION OF THE DIFFERENCE IN THE WIDTHS OF THE TEST- AND INSPECTION-BLOCKS: CONTROL GROUP

Days	Difference in Width of Test- and Inspection-Blocks (Inches)										
	−1.25	−1.00	−0.75	−0.50	−0.25	0.00	0.25	0.50	0.75	1.00	1.25
1 to 3	−0.038	−0.025	−0.069	−0.131	+0.025	−0.063	−0.025	+0.038	+0.031	+0.031	−0.031
4 to 11	+0.063	−0.013	−0.006	−0.025	+0.013	−0.044	−0.031	+0.050	+0.031	+0.019	0.000

Note.—Each entry is the mean difference, in inches, between the postinspection match and the preinspection match. Each mean is based on six *S*s.

For the curve representing the first three days, the means of the first five points and of the last five points are each significantly different from zero ($P < .05$) by the t test, showing that a figural aftereffect did occur.

All of the functions shown in Fig. 2 have the same general shape. The principal effect of repeated satiation seems to be to shift the entire function upward in the coordinate system.

The results for the control group are shown in Table 1. The significance of the difference between the means of the first five entries (narrow T-objects) and the last five entries (wide T-objects) was evaluated by means of the t test. This was done separately for the results of the first three days and the results of the last eight days. In neither case was the difference significant at the .05 level. Thus the control group, as expected, showed no figural aftereffect.

Discussion

It is clear from the results of the experimental group presented in Fig. 1 and 2 that what grows, over the course of 11 days, is a tendency to overestimate *all* test blocks. The fact that in the control group this tendency was either absent (for wide T-blocks) or much reduced (for narrow T-blocks) indicates that it is caused primarily by repeated satiation with the 2½-in. I-block.

What can be said of the behavior of the figural aftereffect over the experimental period depends, of course, on how it is defined. In view of Köhler's generalization that a T-object that is identical with the I-object suffers little or no figural aftereffect, it seems reasonable to define the figural aftereffect obtained as the degree of over- or underestimation of a test object relative to that for the 2½-in. T-object.

Thus the figural aftereffect (FAE) is given by

$$FAE = (T - T_0) - (I - I_0)$$

where T is the matched width of a given test-block after satiation, T_0 the matched width of the same test-block before satiation, I the matched width of the I-block after satiation, I_0 the matched width of the I-block before satiation. In these terms, the figural aftereffect showed virtually no change over the 11-day period.

It should be noted that there is evidence that the amount of figural aftereffect does increase with time of stimulation when the range of times involved is but a few minutes (Charles & Duncan, 1959). The present results have no bearing on this problem since long satiation periods, that probably resulted in maximal effects, were used on each day.

On the basis of the findings presented above it seems quite likely that the alleged long-term growth of the figural aftereffect in kinesthesis does not occur, but that the previous investigators mistook the tendency to overestimate all T-objects for a figural aftereffect. This is possible because these investigators studied only a stimulus situation in which the form of the figural aftereffect was an overestimation of the test-block.

On the other hand, the fact that both Köhler and Dinnerstein (1947) and Wertheimer and Leventhal (1958) satiated both hands in their experiments raises some difficulties for this interpretation of their results. It would seem at first thought that, since both hands were satiated, the scale as well as the test-blocks should be overestimated so that there would be no measurable effect. Such an argument assumes, however, that the two hands received exactly the same treatment. This is not so. The most obvious difference is that the I-blocks presented to the two hands were of different width and may well have had quantitatively different effects. There are a number of other differences that might have played a role in producing the results in question, but the effects of many of the relevant variables are so poorly understood that there is

little point in guessing. Further research is needed to clarify this matter.

Summary

For six *Ss* a figural aftereffect in kinesthesis was measured daily for a period of 11 days. On each day the effect of a single inspection-object upon each of 11 different test-objects was measured. The I-object was a rectangular block of wood 2½ in. wide, the T-objects were blocks varying in width from 0.5 in. to 4 in.

There is no figural aftereffect when the I- and T-objects are approximately equal in width. The widths of T-objects that are narrower than the I-object are underestimated, the widths of T-objects wider than the I-object are overestimated. The degree of over- or underestimation, after increasing rapidly as the difference in the width of the I- and T-objects increases from zero, falls off again for very large differences.

Over the 11-day period of the experiment *Ss* steadily increased their estimate of the width of all the T-objects, but the figural aftereffect, defined in terms of the difference between the estimated widths of the I- and T-objects, showed no change in size.

References

Charles, J. P., & Duncan, C. P. The distance gradient in kinesthetic figural aftereffect. *J. exp. Psychol.*, 1959, **57**, 164–170.

Köhler, W., & Dinnerstein, D. Figural aftereffects in kinesthesis. In *Miscellanea psychologica Albert Michotte*. Paris: Librairie Philosophique, 1947. Pp. 196–220.

Wertheimer, M., & Leventhal, C. M. "Permanent" satiation phenomena with kinesthetic figural aftereffects. *J. exp. Psychol.*, 1958, **55**, 255–257.

(Received December 17, 1959)

Journal of Experimental Psychology
1961, Vol. 61, No. 1, 57–66

THE LOCI OF REINFORCEMENT [1]

GEORGE COLLIER AND LEONHARD MYERS

University of Missouri

In a study examining the relation between concentration of sucrose used as a reinforcer and interreinforcement interval, Collier and Siskel (1959) suggest that the momentary rate of responding is jointly determined by the intensity of the reinforcing stimulus and the momentary postingestive concentration.

Over the course of an experimental session a given combination of concentration of reinforcement, volume per reinforcement, and interreinforcement interval will produce a momentarily changing combination of intensity of stimulation and volume and concentration of postingestive load. This should, by the Collier and Siskel conjecture, produce a momentarily changing rate of responding. Different combinations of the concentration, volume, and interval between reinforcement will produce differential rates of change of reinforcement intensity and postingestive load and thus should result in differential rates of change of responding. The present set of experiments extends the original Collier-Siskel study and explores the effect of these variables on the course of the rate of responding within a session.[2]

METHOD

Apparatus.—Four Skinner boxes delivering liquid reinforcements were used. The magazines were 12-in.-diameter aluminum plates $\frac{1}{8}$ in. thick with 72 $\frac{5}{16}$-in.-diameter cups drilled around the periphery. The maximum volume of each cup was .4 ml. A second aluminum plate on top of the first covered the cups and reduced evaporation. The plate was advanced by a motor-driven Geneva movement. The operation of the magazine was barely audible. The magazine was located below the rat cage, which was a $8 \times 9\frac{1}{4} \times 7$-in. Hoeltge HB-11A cage of the same sort as used to house *S*. A Gerbrands bar requiring a force of approximately 22 gm. located $3\frac{1}{4}$ in. above the magazine was used. The whole assembly was located in a sound-deadened, light-tight, dimly illuminated, ventilated icebox hull. The magazines were loaded by means of a Cornwall automatic pipetting device. The sucrose solutions were prepared from commercial sugar 24 hr. before use and maintained at room temperature for a maximum of 72 hr.

Subjects.—The 96 *S*s of Exp. 1 were male rats, 141–150 days old, of the Sprague-Dawley strain obtained from Holtzman Company. They had been used in a previous experiment on the same apparatus using light-contingent reinforcement. Twenty *S*s were lost during magazine training[3] and four were randomly discarded to equalize the number in each group.

The 24 *S*s of Exp. 2 were female rats, 120–150 days old, of the Sprague-Dawley strain obtained from Holtzman Company. They had been used in a previous experiment with sucrose reinforcement. These same 24 *S*s were used in Exp. 3.

Procedure.—Experiment 1 is essentially a repetition of the Collier and Siskel (1959) experiment, but with a wider range of sucrose concentrations and with independent groups to obviate shift effects. Six concentrations,

[1] This investigation was supported in part by Research Grants M-817 and M-3328 from the National Institutes of Health, Bethesda, Maryland. The computation in the analysis was performed by the University of Missouri IBM center and was supported by the Faculty Research Council. Experiment 1 was performed by the junior author as a yet-to-be-completed MA thesis.

[2] These conjectures as well as brief report of the present experiments were presented by the senior author as a symposium paper, "Interaction of Factors Governing Amount of Reinforcement Function," Midwestern Psychological Association, Detroit, May 1958.

[3] This was our first experience with this magazine and procedures insuring a low loss rate during magazine training had not been developed.

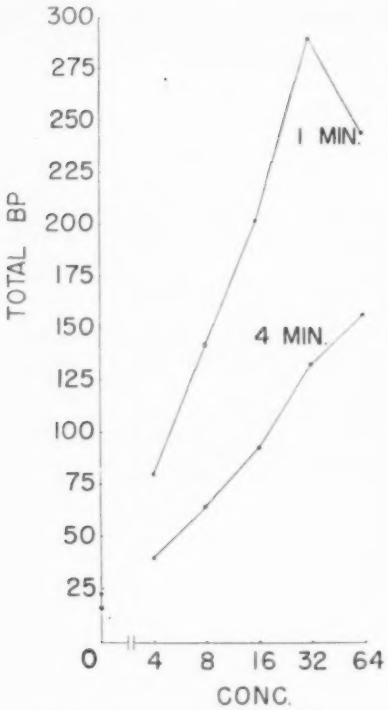

FIG. 1. The total number of bar presses in Exp. 1 as a function of log concentration and interreinforcement interval.

0%, 4%, 8%, 16%, 32%, and 64% by weight, were combined factorially with two fixed interval (FI) schedules, 1 and 4 min. Six days of deprivation accommodation (23 hr. food and water), four days of magazine training (22 hr. food and water), two days of accommodation (22 hr. food, ad lib. water), and 12 test days (22 hr. food, ad lib. water) were given. During magazine training the magazines were operated on a 1-min. variable interval (VI) schedule, each *S* being offered 18 .3-ml. water reinforcements during the 20-min. session. It was possible during magazine training for *S* to be pinched by the magazine plate sliding past the cover. All *Ss* who showed signs of having been pinched or who showed a short run of not taking reinforcements after having started taking reinforcements were discarded. All *Ss* used missed no more than one reinforcement over the last two days. Six *Ss* were assigned at random to each of the 12 groups (six concentrations by two intervals). At the end of the first six days the 1-min. group was shifted to 4 min. and 4 min. to 1 min. for an additional six days of testing. The concentration remained the same: .1 ml. of solution was used under all test conditions. Concentration and intervals were counterbalanced across running times and boxes. The experimental periods were 20 min.

In Exp. 2 a different range of volume and concentration was explored in an investigation of the extent to which volume and concentration act equivalently. Two concentrations, 4% and 32%, were combined with three volumes, .03, .1, and .3 ml. Reinforcements were delivered on a 1-min. FI schedule. Four *Ss* were run in each group. Each combination of variables contained equal numbers from each of the groups in a previous experiment matched on the basis of previous level of responding. Magazine training was not necessary. The *Ss* were run for six days. Experimental periods were 30 min. long. Conditions, boxes, and order of running were counterbalanced. Animals were fed for 1 hr., ½ hr. after running.

Experiment 3 dealt with the question of whether volume behaves the same way as concentration with respect to interreinforcement interval. Three volumes, .03, .1, and .3 ml. were combined factorially with two FI schedules, 1 and 4 min. All groups were run at 32% sucrose for the first five days and 64% for the second five days. The four *Ss* in each group came from the preceding experiment. The *Ss* were run at the same

TABLE 1

ANALYSIS OF VARIANCE ON TOTAL NUMBER OF BP IN TERMINAL SESSION AT EACH INTERVAL IN EXP. 1

Source	*df*	*MS*	*F*
Between *Ss*	71		
Concentration (C)	5	140,022.4	62.40**
I × O	1	1.4	.00
I × O × C	5	1,775.8	.79
Error (b)	60	2,243.9	
Within *Ss*	72		
Interval (I)	1	224,518.0	157.81**
Order (O)	1	9,900.3	6.96*
I × C	5	16,307.5	11.46**
O × C	5	1,721.4	1.21
Error (w)	60	1,422.7	

*$P < .05$.
**$P < .01$.

volume as in the preceding experiment. Equal numbers of *Ss* matched on level of responding from each concentration group were assigned to the two interval conditions. Conditions, boxes, and order of running were counterbalanced. Animals were fed for 1 hr., ½ hr. after running.

Results

Experiment 1.—Figure 1 presents the curves for total number of bar presses (BP) as a function of concentration and interreinforcement interval averaged over the sixth and twelfth days. An analysis of variance of the total number of BP in the 20-min. test session for the sixth and twelfth test day for concentrations of 0%, 4%, 8%, 16%, 32%, and 64% (C), Intervals of 1 and 4 min. (I), and first and second Orders (O), is presented in Table 1. The main effects of C, I, O, and the I × C interaction prove to be significant.

TABLE 2

Percentage of Reinforcements Taken at Each Concentration and Interval in Exp. 1

FI (Min.)	Concentration					
	0	4	8	16	32	64
1	54.8	94.4	100.0	100.0	100.0	99.6
4	68.1	95.8	98.6	100.0	98.6	98.6

The total number of BP in 20 min. is an increasing function of concentration up to 32% sucrose for the 1-min. FI group and an increasing function of concentration up to 64% for the 4-min. FI group. The slope of the rate-concentration function over the lower concentrations also differs as a function of the interreinforcement interval. The Order effect resulted from the fact that number of BP at the end of the

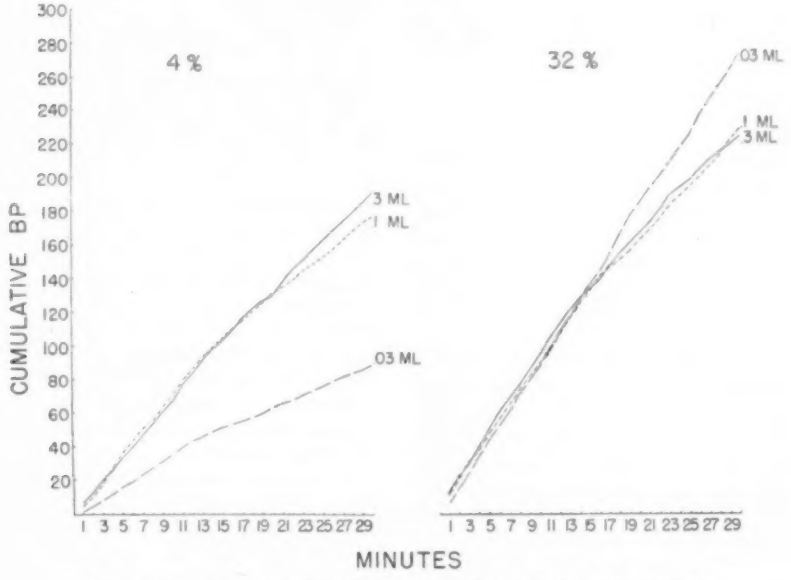

Fig. 2. The cumulative number of bar presses in Exp. 2 as a function of concentration and volume.

TABLE 3

ANALYSIS OF VARIANCE ON BP DATA OF EXP. 2

Source	*df*	*MS*	*F*
Between *Ss*	23		
Volume (V)	2	210.4	.97
Concentration (C)	1	4,295.3	19.78***
V × C	2	802.7	3.70*
Error (b)	18	217.1	
Within *Ss*	1,416		
Minutes (M)	29	55.0	2.27***
Test days (T)	1	26.4	0.38
M × T	29	39.6	1.73*
V × M	58	41.3	1.71**
V × T	2	129.8	1.84
C × M	29	42.0	1.74**
C × T	1	97.7	1.39
C × V × M	58	49.1	2.03***
C × V × T	2	84.4	1.20
V × M × T	58	44.5	1.95**
C × M × T	29	44.2	1.93***
C × V × M × T	58	45.9	2.01***
Error (w)	1,062		
e_1	522	24.2	
e_2	18	70.4	
e_3	522	22.8	

Note.—The error terms for the within analysis follow the following pattern: M, VM, CM, CVM against Error$_1$, T, VT, . . . against Error$_2$ and MT, ... CVMT against Error$_3$.

* $P < .05$.
** $P < .025$.
*** $P < .01$.

second six days (134.4) was greater than at the end of the first six (115.9).

The percentage of reinforcements taken for each combination of interval and concentration is presented in Table 2. Only the 0% group showed any tendency not to take all of the reinforcements offered.

An analysis of variance similar to that in Table 1, on the mean duration of the BP, determined by dividing the total time the bar was depressed in a session by the total number of BP, showed no significant sources of variation. There was a tendency for the duration of the BP to be slightly longer (.82 vs. .91 sec.) for the 4-min. FI condition, $P < .10$ ($F = 3.05$; $df = 1/60$).

An analysis of variance on the weight of the rats on the sixth and twelfth day similar to that presented in Table 1 showed only Order to be significant (first = 346 gm.; second = 355 gm.). This may result from the cumulative effects of the amount of sugar consumed.

Experiment 2.—In Fig. 2 are presented the cumulative BP curves by 1 min. over 30-min. sessions of Exp. 2. The points are the average of the last two test days at each combination of the variables. An analysis of variance of the data (in the noncumulative form) for Concentrations of 4% and 32% (C), Volumes of .03, .1, .3 ml. (V), 1-min. intervals (M), and last two Test days (T) is presented in Table 3. Here the main effects of C and M were significant. The failure of the main effect of V to be significant results from the interactions with C and M discussed below.

The total number of responses is greater at 32% than at 4%. This is true for the initial rate (4%: 4.8 vs. 32%: 10.7 BP in Minute 1) as well as the total number (4%: 151.8 vs. 32%: 255.5 BP). In the first minute of responding the number of responses is an increasing function of the volume offered at each concentration. On the other hand, the total number of responses in the 30-min. session is an increasing function of the volume offered at 4% and a decreasing function of the volume offered at 32%, yielding the significant C × V interaction not found in the following experiment which compared only 32% and 64% solutions.

As can be seen from the data of Exp. 3, the volume differences in the first minute at 32% are not very stable ones and are probably not significant. The order of differences obtained between the combinations

of volume and concentration are the result of both the initial rate differences and differential declines in rate. At the 4% concentration, the rate of decline is an inverse function of volume (i.e., the greatest decline occurs for the smallest volume) while at the 32% concentration, the rate of decline is a direct function of volume, yielding the significant V×M interaction not found in the following experiment. It appears that a within-session decline in rate of responding can arise either with combinations of low concentration and low volume or with high concentration and high volume.

The interactions of the major variables with sessions is again the result of shift effects. While the order of the differences is the same, the magnitudes are changing. In Fig. 3 are presented the daily total number of responses for each group. It shows, after an initial raise, the typical decline in rate over days.

Experiment 3.—In Fig. 4 are presented the cumulative BP curves by 1 min. over 30-min. sessions of Exp. 3. The points are the average of the last two days at each combination of the variables. An analysis of variance of these data in the non-cumulative form for Concentrations of 32% and 64% (C), Volumes of .03, .1, and .3 (V), interreinforcement intervals of 1 and 4 min. (I), 1-min. intervals (M), and last two test days (T) is presented in Table 4. The main effect of M and a number of the interactions were significant. Here the failure of the main effects of either C or V to be significant results from their interaction with each other and with I (see below).

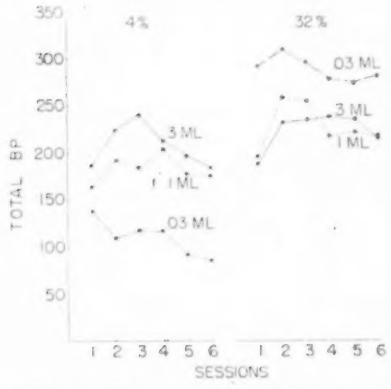

FIG. 3. The total number of bar presses in Exp. 2 over sessions as a function of concentration and volume.

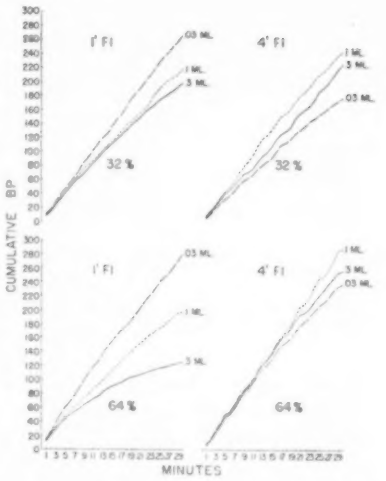

FIG. 4. The cumulative number of bar presses in Exp. 3 as a function of concentration, volume, and interval.

Under both concentrations (32% and 64%) on the 1-min. schedule the highest rate and the largest total number of responses occurs for .03 ml., the lowest and smallest for .3 ml. For the 4-min. schedule, on the other hand, the highest rate and total are for .1 ml. and the lowest and smallest for .03 ml. In both cases (1 and 4 min. FI), the highest momentary rates occur for 64% at all volumes,

TABLE 4

ANALYSIS OF VARIANCE ON BP DATA OF EXP. 3

Source	*df*	*MS*	*F*
Between *Ss*	(23)	785.1	
Volume (V)	2	549.1	0.74
Interval (I)	1	353.5	0.47
I × V	2	1,595.7	2.14
Error (b)	18	745.3	
Within *Ss*	(2,856)	19.8	
Concentration (C)	1	36.2	0.25
Minutes (M)	29	128.9	7.58***
Test Days (T)	1	30.6	0.66
C × M	29	35.0	2.46***
C × T	1	6.0	0.23
M × T	29	16.3	1.18
V × C	2	214.4	1.49
C × I	1	1,169.2	8.14**
V × M	58	12.5	0.74
I × M	29	142.3	8.36***
V × T	2	13.3	0.29
I × T	1	79.0	1.69
I × V × C	2	59.9	0.42
I × V × M	58	21.9	1.29
I × V × T	2	103.6	2.21
I × C × M	29	17.4	1.22
I × C × T	1	34.5	1.33
I × M × T	29	19.6	1.42
V × C × M	58	20.2	1.49**
V × C × T	2	55.6	2.14
V × M × T	58	11.0	0.80
C × M × T	29	22.4	1.50*
I × V × C × M	58	16.0	1.13
I × V × C × T	2	4.5	0.17
I × V × M × T	58	14.6	1.06
I × C × M × T	29	25.2	1.69*
V × C × M × T	58	13.8	0.92
I×V×C×M×T	58	15.3	1.03
Error (w)	(142)	16.4	
e_1	18	143.7	
e_2	522	17.0	
e_3	18	46.8	
e_4	522	14.2	
e_5	18	26.0	
e_6	522	13.8	
e_7	522	14.9	

Note.—The error terms for the within analysis follow the following pattern: C, IC, VC, IVC against e_1; M, IM, ... IVM against e_2, ... and CMT, ICMT, VCMT, and IVCMT against e_7.

* $P < .05$.
** $P < .025$.
*** $P < .01$.

as can be seen by comparing the slopes of the cumulative curves. The rate of decline in rate of responding is a joint function of concentration and volume. The most rapid declines in rates occur for high concentration at large volumes and short intervals, and at low concentration, small volumes and long intervals.

The highest rates and total numbers of responses achieved in either Exp. 2 or 3 are, interestingly enough, for 64% sucrose at either .03 ml. or at 4-min. interreinforcement interval. These results argue against any interpretation of the volume-concentration effects in terms of dilution of the reinforcement by saliva or of the interval effects in terms of the number of reinforcements received. The C×I interaction results from the fact that the total number of responses is a decreasing function of concentration at 1-min. interreinforcement interval (32%: 231.9 and 64%: 200.4), while the total is an increasing function at the 4-min. interval (32%: 214.7 and 64%: 259.6). The rate of decline is also a function of the interreinforcement interval (I × M), the greatest rate resulting from the shortest interval. Both the amount consumed and the rate of delivery are important

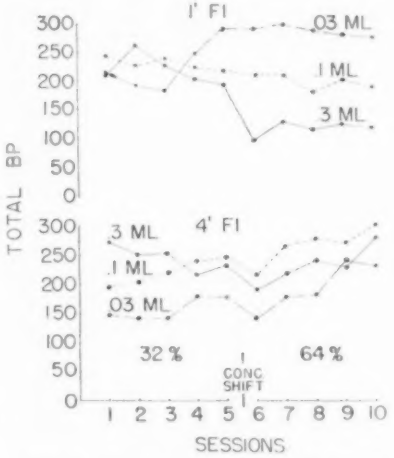

FIG. 5. The total number of bar presses in Exp. 3 over sessions as a function of concentration, volume, and interval.

determinants, as can be seen from Fig. 4.

The interaction of the major variables with sessions is again the result of shift effects. Figure 5 presents the daily totals for each group. It is apparent that a number of days is required to reach a stable rate in responding following a shift in concentration or volume of reinforcement.

In summary of the preceding three experiments, the following relations can be asserted:

1. The *initial rate* of responding is an increasing function of the concentration of sucrose offered. The increment in rate for equal (log) increments in concentration is a decreasing function of the interreinforcement interval and an increasing then decreasing function of the volume (see Collier and Willis, 1960, and Table 5) of the reinforcement offered. Shifts in the parameters of amount result in gradual rather than immediate shifts in rate of responding over sessions.

2. The *initial rate* of responding is an increasing function of the volume of sucrose offered. The increment in rate for equal (log) increments in volume is a decreasing function of the concentration of the reinforcement offered. The initial-rate-volume as well as the total BP-volume function becomes nonmonotonic for "high" concentrations of sucrose (see Table 5).

3. The *initial rate* of responding is a decreasing function of interreinforcement interval at low concentration or small volumes and an increasing function of interval, over the range of intervals tested, at high concentration or large volumes. The highest initial rates obtain in the present experiment for the combinations of long interreinforcement intervals at high concentration.

TABLE 5

NUMBER OF RESPONSES IN THE FIRST MINUTE OF RESPONDING FOR EACH COMBINATION OF VOLUME AND CONCENTRATION IN EXP. 2 AND 3

Concentration[a]	Volume (ml.)		
	.03	.1	.3
4%	1.8	5.4	7.1
32%	8.6	10.9	11.4
64%	13.5	11.6	12.0

[a] The 4% values are from Exp. 2, the 32% values from Exp. 2 and 3, and the 64% values from Exp. 3. All values are for 1-min. FI.

4. The *momentary rate* of responding is a decreasing function of number of reinforcements. The rate of decline is itself a joint increasing function of the concentration and volume of the reinforcement, and a decreasing function of the interreinforcement interval.

DISCUSSION

The occasion of food or water reinforcement can be partitioned into a chain of events: the distal stimuli and the responses to them; the proximal stimuli and the responses to them; the consummatory responses and the stimuli resulting from them; the postingestive processes; and finally, changes in nutritional state. The events associated with reinforcement have been assumed to be located at various points along this chain or even beyond it in some cases.

Most attempts to determine the *locus* of reinforcement have involved either the interruption of the chain, e.g., the esophageal fistula or latent learning experiments, or bypassing a portion of the chain, e.g., the stomach loading or injection experiments. An alternative approach, the present one, is to examine the intact chain quantitatively in terms of the parameters of reinforcement (e.g., time, amount, kind, etc). An examination of the network of relations found will determine whether a single process or more than one is required, and place

TABLE 6

AVERAGE LATENCY OF THE FIRST RESPONSE FOLLOWING A REINFORCEMENT AND AVERAGE MOMENTARY RATE FOLLOWING THE FIRST RESPONSE: EXP. 2

Volume (ml.)	Concentration (%)	Latency (sec.)	BP/Sec.
.03	4	38.8	.10
.1	4	34.1	.25
.3	4	33.0	.23
.03	32	28.6	.24
.1	32	31.7	.29
.3	32	35.3	.28

some empirical constraints on the mechanisms of reinforcement considered. On the basis of such an examination, Collier and Siskel (1959) have offered a simple account of food reinforcement which envisages at least three independent loci of events controlling momentary rate of performance, the proximal reinforcing stimuli, the momentary postingestive concentration, and the nutritive state of *S*.

Initial Rate

Since the lumbering processes of digestion and absorption neither allow the mediation of initial rate by concurrent postingestive events, nor preserve the quantitative, qualitative, and temporal information contained in the parameters of amount, it appears that the burden of accounting for the initial-rate-amount functions falls upon the immediate consequences, olfactory, gustatory, motor, etc., of ingestion.

The simplest assumption that accounts for the relations observed in amount studies is that initial rate of instrumental responding increases linearly with the log intensity of the proximal reinforcing stimuli. The slope, intercept (i.e., the rate at threshold intensity), and the point of slope change or maximum of this function are functions of variables such as volume, interreinforcement interval, quality, deprivation, etc.

Volume.—Initial rate is an increasing function of volume. Over the lower ranges of volume and concentration there is reciprocity of their effects on initial rate. Thus, the rate-volume function (at low concentrations) and the rate-concentration function (at small volumes) have the same slope (Table 5). But, the effects of volume and concentration on rate are not additive. Over the lower range of concentrations the slope and the intercept of the rate-concentration function are increasing functions of volume (Collier & Willis, 1960) forming a family of diverging rate-concentration curves, while at the upper ranges the slope is a decreasing function of volume (Table 5).

Two sets of events are determined by increases in volume per reinforcement. First, beyond a certain minimal value, increases in volume lead to increases in the number of licks per reinforcement, each lick having the stimulus intensity associated with the concentration and taste of the ingested substance. Second, increases in volume lead, through the increased number of licks, to an increase in the number of receptors sampled, and to repeated stimulation of some. Because of overlap, the number of new receptors sampled per reinforcement decreases as number of licks increases, which might account for the nonlinear character of the rate-log volume function. The extent to which the effects of volume are separately or jointly products of these two sets of events is difficult to determine. It is conceivable that the result of successive stimulations at the same intensity is the basic parameter of volume per reinforcement, but no direct measures are extant. Hellyer (1953) has shown that duration of consummatory activity, per se, influences running speed. On the other hand, Knarr (1959) has shown that when a taste factor (concentration of sucrose) is introduced, it far outweighs both the duration and amount of activity factors. The latency measures in Table 6 indicate that consummatory activity does not differentially compete with BP as a function of volume over the range of volumes considered here.

Some intimation of the mode of operation of volume and concentration on rate can be obtained from Table 6, which exhibits the momentary rate and latency of BP in Exp. 2. Latency is defined as the time from the last reinforcement to the first response in the interval between reinforcements. The momentary rate is the ratio of the number of responses to the time from the first response in the interval to the next reinforcement.[4] Three results, borne out by a later, more elaborate study (Collier, 1959) are pertinent here. Latency increases over time while momentary rate, except for an initial warm-up, does not. That is, the changes in rate of BP over the duration of a session are the result of increased latencies rather than changes in the rate of BP (cf. Stellar & Hill, 1952). The rate-volume relation is nonmonotonic, contrary to the data with latency included presented earlier. The subsequent study (Collier, 1959) shows that the same is true for the rate-concentration function at high volumes. These two findings suggest that volume per reinforcement is the key factor in producing inversions and indicate that the initial supposition of Collier and Siskel (1959) that the "aversion limb" of these functions is entirely a function of postingestive effects is incorrect. Thus, rate is first an increasing then a decreasing function of volume. The point at which the inflection in slope occurs is a function of concentration.

Interreinforcement interval.—The slope and the intercept of the initial-rate concentration function are both decreasing functions of interval, while the point of inflection of slope is an increasing one. Thus, varying interval produces a family of diverging rate-concentration curves with maxima at different points. These relations, both the decremental and incremental, suggest a summative process. The times involved are too long for it to be attributed to the receptor processes.

Both quality (Collier, 1959) and deprivation (Collier & Willis, 1960) have been examined for their effect on the parameters of the initial rate-concentration function. Aside from determining the threshold concentration, quality influences the slope. Equal increments in log concentration do not appear to produce equal increments in rate, as would be expected on the simplest rate-intensity hypothesis. The slope of the rate-concentration function is an increasing function of deprivation for food-deprived *Ss* (Collier & Willis, 1960). The slope becomes negative for water-deprived *Ss*. These experiments suggest that the intensity of the proximal reinforcing stimuli is a joint product of concentration and deprivation.

Terminal Rate

Within-session declines in rates of responding as a function of time or number of reinforcements have often been shown. Three common accounts of such declines are satiation, response habituation, and extinction. The first is attributed to the occurrence of the reinforcement, the second to the occurrence of the response, and the third to the occurrence of the response without reinforcement. The rate decrements resulting from satiation and habituation dissipate in time while those resulting from extinction, within the limits of spontaneous recovery, do not. Within-session declines in rates of responding occur in the present experiments under two sets of circumstances: combinations of large volume, high concentration, and short intervals and combinations of small volume, low concentration, and long intervals. The first, occurring as a function of the amount and character of the load, suggests satiation. The second, occurring as a function of time under conditions of minimal reinforcement, suggests response habituation.

The results of the present experiments show that when the rate of delivery is controlled so that the limit of solution

[4] Momentary rate differs from local rate in that it is the average rate of responding and does not represent the density of responding in any "bursts" of responding which occur.

that the animal will ingest during the session is not reached, rate of responding is an increasing function of concentration and the preference-aversion point, i.e., the maximum of the rate-concentration function becomes a function of concentration, rate of delivery, length of session (number of reinforcements), and volume of reinforcement. The form of the preference-aversion function under conditions of free ingestion appears to be a result of the confounding of ingestive and postingestive factors.

These facts indicate that "momentary satiation" is a function of momentary postingestive concentration and therefore of the colligative properties of the ingested substance. The rapidity with which it can occur further strengthens the view that momentary satiation is independent of any change in the nutritive state of the organism or of any of the associative metabolic effects such as bodily heat, as proposed by Strominger and Brobeck (1953). The studies of Kraukauer and Dallenbach (1937) indicate that it is not due to adaptation of the gustatory receptors since the rapidity of recovery from adaptation is outside the limits required to account for "satiation" in the present experiment.

The occurrence of a within-session decline with a between-sessions recovery (Fig. 3 and 5) under conditions of minimal reinforcement suggests that the occurrence of a response leads to a lowered probability of its occurrence, as long ago suggested by Hull (1943), that this decrement in probability dissipates in time, and that the decrement is counterbalanced by reinforcement. More dramatic within-session declines have been demonstrated in situations where even more minimal reinforcements, such as light onset are used, and it has been shown that the amount of recovery is a function of the duration of the between-session intervals (Premack, Collier, & Roberts, 1957).

Summary

On the basis of variations of concentration, volume, and the interval between sucrose reinforcements in a bar-pressing situation with rats as *Ss*, it was shown that initial rate of responding is an increasing function of concentration and volume and a decreasing function of interval, while the terminal rate is a decreasing function of concentration and volume and an increasing function of interval. Analysis of these results led to the conclusion that two independent processes govern rate of responding in this situation, taste and momentary satiation. The former reflects sensory intensity and quality, the latter, the momentary postingestive concentration.

References

Collier, G. The loci of reinforcement. *Amer. Psychologist*, 1959, **14**, 398. (Abstract)

Collier, G., & Siskel, M., Jr. Performance as a joint function of amount of reinforcement and inter-reinforcement interval. *J. exp. Psychol.*, 1959, **57**, 115–120.

Collier, G., & Willis, F., Jr. Regulation and reinforcement. Paper read at Midwest. Psychol. Ass., St. Louis, April 1960.

Hellyer, S. Duration of the consummatory response as a variable in amount of reinforcement studies. Unpublished doctoral dissertation, Indiana Univer., 1953.

Hull, C. L. *Principles of behavior*. New York: Appleton-Century, 1943.

Knarr, F. A. Performance as a function of differential consummatory activity and reinforcement. Unpublished master's thesis, Univer. Missouri, 1959.

Kraukauer, D., & Dallenbach, K. M. Gustatory adaptation to sweet, sour, and bitter. *Amer. J. Psychol.*, 1937, **49**, 469–475.

Premack, D., Collier, G., & Roberts, C. L. Frequency of light contingent bar pressing as a function of the amount of deprivation for light. *Amer. Psychologist*, 1957, **12**, 411. (Abstract)

Stellar, E., & Hill, J. H. The rat's rate of drinking as a function of water deprivation. *J. comp. physiol. Psychol.*, 1952, **45**, 96–102.

Strominger, J. L., & Brobeck, J. R. A mechanism of regulation of food intake. *Yale J. Biol. Med.*, 1953, **25**, 383–389.

(Received December 21, 1959)

Journal of Experimental Psychology
1961, Vol. 61, No. 1, 67–75

THE EFFECTS OF AWARENESS ON VERBAL CONDITIONING [1]

SAUL M. LEVIN [2]

Duke University

A number of experiments have demonstrated that interpersonal reinforcers can be employed to condition a variety of verbal responses (e.g., Cohen, Kalish, Thurston, & Cohen, 1954; Greenspoon, 1955; Verplanck, 1955). In most of these studies it has been inferred, on the basis of *S*'s response to a postexperimental interview, that learning occurred without *S*'s awareness of a correct contingency between the reinforced response and the reinforcer. Much of the evidence for conditioning without awareness comes from studies in which a sentence construction task developed by Taffel (1955) has been employed as the conditioning procedure (Buss, Gerjuoy, & Zusman, 1958; Cohen et al., 1954; Essman, 1959; Hartman, 1955; Kirman, 1958; Klein, 1954; Leventhal, 1959). In this procedure, *E* selectively reinforces sentences beginning with certain pronouns. Upon completion of the experimental task an interview for awareness is conducted. The validity of the evidence for conditioning without awareness would seem to be partially dependent upon the sensitivity of the postexperimental interviewing procedures.

In many verbal conditioning studies the postexperimental interviews have been brief and have consisted of general questions. Cushing (1957), employing a conditioning technique other than the sentence construction task, has reported evidence which suggests that aware *Ss* may not verbalize their awareness in responding to such an interview. None of Cushing's *Ss* verbalized awareness in responding to the first two questions of a seven-question postexperimental interview. However, 27 *Ss* verbalized awareness of a correct contingency in responding to subsequent questions. The first two questions in Cushing's interview were nonspecific in nature and were similar to the questions typically employed to investigate awareness. The remaining questions were more specific and relevant to the contingency.

Another factor which might reduce the sensitivity of a postexperimental interview is the practice of questioning *Ss* following extinction trials (Buss et al., 1958; Cohen et al., 1954; Dailey, 1953; Essman, 1959; Grossberg, 1956; Hartman, 1955; Klein, 1954). An *S* who may have been aware of a correct contingency during the conditioning trials might not have reported it during the typical brief interview either because he forgot it or because he got the idea, during the extinction trials, that he had been wrong.

The primary purpose of the present study was to evaluate the extent to which previous evidence for conditioning without awareness is an artifact of insensitive interviewing pro-

[1] This study is based on a dissertation submitted to the Psychology Department of Duke University in partial fulfillment of the requirements for a PhD in Psychology. The author wishes to thank C. D. Spielberger, director of the dissertation, for his encouragement and guidance.

[2] Now at Lafayette clinic and Wayne State University.

cedures. A sentence construction task was employed as the conditioning technique and *Ss* were interviewed immediately following the acquisition trials according to a detailed interview schedule.

METHOD

Subjects.—The *Ss* were 79 white male patients at the Durham VA Hospital, 50 of whom were hospitalized for neuropsychiatric conditions and 29 for medical disorders. Patients with diagnoses of organic brain damage were not included in the sample. Only neuropsychiatric patients on open wards and medical patients not suffering from acute physical illness served as *Ss*.

Stimulus materials.—The stimulus materials for the conditioning task consisted of 100 3-in. × 5-in. white index cards. A different past tense verb, selected from a list of the 1,000 most frequent words in written English (Thorndike & Lorge, 1952), was typed on each card. The pronouns *I, we, you, he, she, they* were typed in a different sequence below the verb on each card.

Procedure.—The experiment was conducted in two sessions on consecutive days. For Session I *Ss* were seen either individually or in groups with not more than 5 *Ss* in a group. At the beginning of the session *Ss* were told that the purpose of the experiment was to find out about the feelings and attitudes of hospital patients. The Ohio Literacy Test (Foster & Goddard, 1924) was then administered and *Ss* with scores corresponding to reading levels below the sixth grade were excused from further participation. Two personality scales, the Taylor Manifest Anxiety Scale and the Edwards Personal Preference Schedule, were then given.[1]

Prior to Session II, *Ss* were assigned to Experimental (reinforced) and Control (nonreinforced) groups equated as closely as possible with respect to age, reading level, psychiatric diagnosis, drug status, and ratio of neuropsychiatric to medical patients. There were no statistically significant differences between the groups on any of these variables. Sixty *Ss* were assigned to the Experimental group and 19 *Ss* to the Control group.

In Session II *Ss* were seen individually for the conditioning procedure. They were instructed to respond to each of the stimulus cards by saying the first sentence which came to mind beginning with one of the pronouns and containing the verb. The cards were then presented one by one in a metal card holder and *S*'s sentences were recorded verbatim. The order of card presentation was the same for all *Ss*. The intertrial interval was 5 to 10 sec. The *E*'s recording activity was shielded from *S* by means of a cardboard screen. The procedure remained the same for all *Ss* for Trials 1–20. For the Experimental group, during Trials 21–100 *E* responded with the word "good" in a flat unemotional tone at the end of any sentence which *S* began with *I* or *we*. For the Control group there was no change from the procedure of Trials 1–20.

Immediately following the conditioning trials each *S* was carefully interviewed in order to elicit information regarding his awareness, during the conditioning trials, of a correct contingency between his behavior and *E*'s reinforcer. The content of the individual interview questions and the sequence of questions were designed to avoid suggesting a correct contingency to *S*. In order to facilitate comparison with previous research, the first four questions of the interview were approximately the same as those used to evaluate awareness in other verbal conditioning experiments (e.g., Cohen et al., 1954; Klein, 1954; Taffel, 1955). Each *S* was interviewed according to the following interview schedule and his responses were recorded verbatim:

1. Did you usually give the first sentence which came to your mind?
2. How did you go about deciding which of the words to use?
3. Did you think you were using some of the words more often than others? Which words? Why?
4. What did you think the purpose of this was?
5. What did you think about while going through the cards?
6. While going through the cards did you think that you were supposed to make up your sentences in any particular way?
7. Did you get the feeling that you were supposed to change the way in which you made up your sentences? How?

(If, in answering Questions 1–7, *S* mentioned the fact that *E* had said "good," Questions 8–10 were not asked since they were designed to investigate *S*'s awareness of the reinforcer.)

8. Were you aware of anything else that went on while you were going through the cards? (If *S* mentioned "good," Questions 9 and 10 were not asked.)
9. Were you aware of anything about me? (If *S* mentioned "good," Question 10 was not asked.)
10. Were you aware that I said anything? (If *S* failed to mention "good" in answering this question, the interview was terminated since the remaining questions all refer to *S*'s reaction to "good." For

[1] Examination of the results revealed that these measures were not related to either awareness or performance on the conditioning task.

Control *Ss* the interview was terminated following this question.)

11. What did my saying "good" mean to you?
12. Did you try to figure out what made me say "good" or why or when I was saying "good"? (If *S* answered "no," Question 13 followed Question 12.)
13. How hard would you say that you tried to figure out what was making me say "good"?
    very hard fairly hard not hard at all
14. What ideas did you have about what was making me say "good"?
15. Would you say that you wanted me to say "good"
    very much? some? didn't care one way or the other?
16. While going through the cards did you think that my saying "good" had anything to do with the words that you chose to begin your sentences? What?

If *S* verbalized a correct contingency at any time during the interview, the above schedule was discontinued and the following questions were asked:

(A) Is that something you were actually aware of while going through the cards or is it something you thought of just now?
(B) Do you remember when, while going through the cards, that idea occurred to you?
(C) Did the fact that you realized this have any effect on the way in which you made up your sentence? In other words, did you try and make up your sentences in that way because I was saying "good"?

All *Ss* who verbalized a correct contingency were also asked Question 15.

At the conclusion of the interview *S* was asked if he had known anything about the experiment beforehand. None of the *Ss* indicated prior knowledge. Each *S* was urged not to communicate with other patients regarding any aspect of the experiment.

Results

Incidence of Awareness

In the present experiment *S* was considered to be aware if, during the postexperimental interview, he stated that "good" had followed sentences beginning with *I* alone, with *we* alone, or with *I* or *we*. These contingencies, if acted upon, would bring a reinforcement 100% of the time. It was further required that *S* indicate, in answering Question A, that he had been aware of the contingency during the conditioning trials. Two bases for inferring awareness were employed: (*a*) *S*'s responses to the first four questions of the postexperimental interview, which approximated the questions used in previous research; (*b*) *S*'s responses to the extended interview.

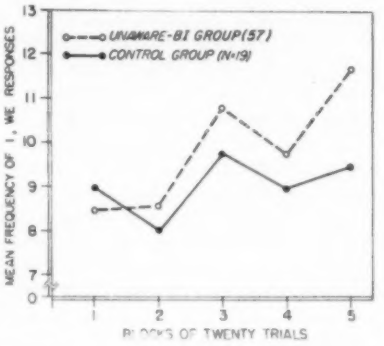

FIG. 1. Mean frequencies of *I, we* responses for Unaware-BI and Control groups.

When responses to the first four interview questions were used as the basis for inferring awareness, only 3 of the 60 Experimental *Ss* were classified as aware. Using the extended interview, 16 additional *Ss* were classified as aware. Therefore, it would appear that a brief interview does not identify all of the aware *Ss*.

Analysis of Conditioning Data

Conditioning without awareness as a function of the criterion of awareness.—Since verbal conditioning studies have usually been concerned with demonstrating conditioning without awareness, the data of *Ss* who verbalized a correct contingency are typically excluded from statistical analysis. Therefore, in order to facilitate comparison with previous findings, the data of the three *Ss* who verbalized an awareness of a correct contingency in responding to the first four interview questions in the present study were not included in any of the following analyses.

The 57 *Ss* who were inferred to be unaware on the basis of their failure to verbalize a correct contingency in responding to the first four interview

questions will henceforth be referred to as the Unaware-Brief Interview (BI) group. The performance of the Unaware-BI group was compared to the performance of the Control group. For this and subsequent analyses the conditioning data were divided into five blocks of 20 trials each. Figure 1 shows the mean frequencies of *I, we* responses over the five blocks of trials for the Unaware-BI and Control groups. These data were analyzed by means of an analysis of variance (Lindquist, 1953, Type I) in which Unaware-BI vs. Control groups was the between-*Ss* factor and Trials was the within-*Ss* factor. In order to control for individual differences in initial operant level of *I, we* responses, difference scores were employed as the response measures for each *S* in the analysis of variance. These measures were derived by subtracting the frequency of sentences beginning with *I* or *we* in the initial block of trials from the frequency of such sentences in each of the four successive blocks of reinforcement trials. The significant effect of Groups in this analysis ($F = 4.11$; $df = 1/74$; $P < .05$), when considered together with Fig. 1, indicates that the Unaware-BI group showed a greater increase of *I, we* responses from its initial operant level than did the Control group. Significant effects of Trials in this and subsequent analyses indicate that for the experimental task there were differing numbers of *I, we* responses elicited by the four reinforcement blocks. This effect might have been partially independent of reinforcement since an analysis of variance (Lindquist, 1953, Treatments by *Ss*) of the difference scores of the Control group alone yielded an effect of Trials which approached significance ($F = 2.57$; $df = 3/54$; $P < .10$). Thus, it would appear that the successive blocks of trials tended to elicit differential frequencies of *I, we* sentences independent of reinforcement.

The extent to which the performance of the Unaware-BI group in the preceding analysis was affected by *Ss* whose awareness was not elicited by the first four interview questions was next evaluated. The conditioning data were analyzed with *S*'s verbalization of a correct contingency subsequent to Question 4 as the criterion of awareness. The 57 *Ss* who constituted the Unaware-BI group in the preceding analysis were divided into an Aware and an Unaware group on the basis of their responses to the extended interview. This Unaware group will henceforth be referred to as the Unaware-Extended Interview (EI) group. The Aware group consisted of the 16 *Ss* who did not verbalize a correct contingency in responding to the first four questions but who did so in responding to the remainder of the interview. The Unaware-EI group consisted of the 41 *Ss* who did not verbalize a correct contingency at any time during the interview. Figure 2 shows the mean frequencies of *I, we* responses over the five blocks of trials for the Aware, Unaware-EI, and Control groups. The difference scores of the three groups were subjected to an analysis of variance (Lindquist, 1953, Type I) in which Aware, Unaware-EI, and Control groups constituted the between-*Ss* factor and Trials was the within-*Ss* factor. This analysis yielded a significant effect of Groups ($F = 5.10$; $df = 2/73$; $P < .01$).

In order to determine the source of the significant effect of Groups in the preceding analysis, three separate analyses of variance (Lindquist, 1953, Type I) were performed comparing

the Unaware-EI and Control, Aware and Control, and Aware and Unaware-EI groups. The effect of Groups was significant in the analyses comparing the Aware group with the Control group ($F=9.71$; $df=1/33$; $P<.005$) and with the Unaware-EI group ($F = 5.79$; $df = 1/55$; $P < .05$), but not in the analysis comparing the Unaware-EI group with the Control group ($F=1.78$; $df=1/58$; $P<.20$). These results, when considered together with Fig. 2, indicate that the Aware group showed a greater increase of *I, we* responses from its initial operant level than did either the Unaware-EI group or the Control group. The latter two groups did not differ from each other.

Specificity of conditioning to the aware pronoun.—Not all of the 16 Aware *S*s verbalized the same contingency; eight stated that the reinforcer had followed sentences beginning with *I* while two said that it had followed sentences beginning with *we*. Six *S*s were aware that *E* had reinforced sentences beginning with *I* and *we*. Since the response measures in the preceding analyses always involved the frequency of sentences beginning with *I* and *we*, two questions arise with respect to the performance of the 10 *S*s who were aware of the reinforcement of only one first-person pronoun: (*a*) Did these *S*s respond differentially to the reinforced pronoun for which they were aware of the reinforcement (Reinforced Pronoun-*S* Aware: RP-A) as compared to the reinforced pronoun for which they were unaware of the reinforcement (Reinforced Pronoun-*S* Unaware: RP-U)? (*b*) Did conditioning without awareness of the RP-U occur?

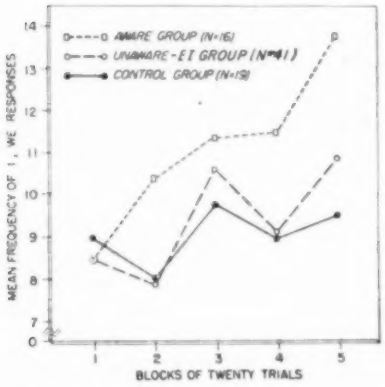

FIG. 2. Mean frequencies of *I, we* responses for Aware, Unaware-EI, and Control groups.

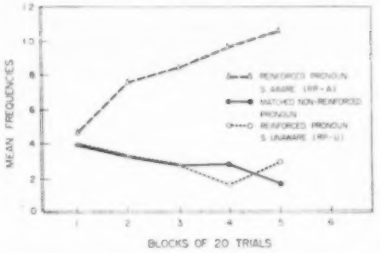

FIG. 3. Mean frequencies of reinforced pronoun: *S* aware, reinforced pronoun; *S* unaware and matched nonreinforced pronoun.

The mean frequencies for the RP-A and the RP-U for the 10 *S*s who were aware of the reinforcement of one pronoun are plotted in Fig. 3. For purposes of statistical analysis, the difference between the number of sentences beginning with the RP-A and the RP-U in each block of trials was computed for each *S*. These data were analyzed by means of an analysis of variance (Lindquist, 1953, Treatments by *S*s) in which the within-*S*s factor was Trials. The significant effect of Trials ($F = 3.00$; $df = 4/36$; $P < .05$), when considered together with Fig. 3, indicates that over the five blocks of trials *S*s tended to give progressively more of the RP-A than the RP-U.

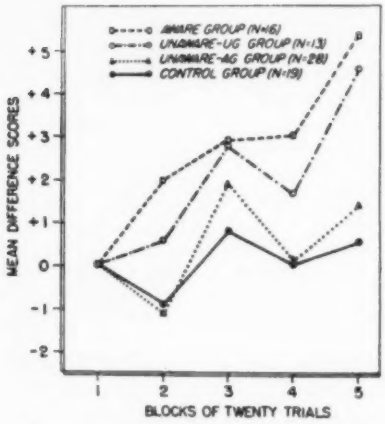

FIG. 4. Mean difference scores for *I, we* pronouns for Aware, Unaware-UG, Unaware-AG, and Control groups.

The finding that *Ss* tended to perform at a higher level on the RP-A than on the RP-U does not rule out the possibility that conditioning without awareness of the RP-U occurred. This possibility was tested by comparing performance on the RP-U with performance on a nonreinforced pronoun which was matched with the RP-U for initial operant level. If conditioning without awareness of the RP-U occurred, then the frequencies for this pronoun should be greater than for the matched nonreinforced pronoun over the reinforcement trials. Figure 3 indicates that *Ss* performed at practically identical levels on the RP-U and the matched nonreinforced pronoun. It would therefore appear that conditioning without awareness of the RP-U did not occur.

Conditioning and awareness of the reinforcer.—In most verbal conditioning studies awareness has referred to verbalization of a correct contingency between a reinforced response and a reinforcer. There are few data available on the performance of *Ss* who do not verbalize an awareness of the reinforcer itself. In the present study, of the 41 *Ss* who were unaware of a correct contingency (Unaware-EI group), 13 failed to mention *E*'s "good" in answering Question 10 of the postexperimental interview. Ten of these *Ss* were unaware that *E* had said anything while the remaining three *Ss* incorrectly identified what *E* had said. These 13 *Ss* who, in addition to being unaware of a correct contingency were also unaware of "good," will be designated hereafter as the Unaware-UG group. The remaining 28 Unaware *Ss* who were unaware of a correct contingency but aware of the "good" will be designated as the Unaware-AG group.

The mean *difference scores* for the Aware group, the Unaware-AG group, the Unaware-UG group, and the Control group are plotted in Fig. 4. Three analyses of variance (Lindquist, 1953, Type I) were performed in which the difference scores of the Unaware-UG group were compared, in turn, with those of the Control group, the Unaware-AG group, and the Aware group. A significant effect of Groups in the comparison between the Unaware-UG and Control groups ($F = 5.03$; $df = 1/30$; $P < .05$) and a nonsignificant effect of Groups in the comparison between the Unaware-UG and Aware groups ($F < 1$) indicates that the Unaware-UG group showed a greater increase in *I, we* responses than the Control group and as great an increase as the Aware group. The effect of Groups in the comparison between the Unaware-UG group and the Unaware-AG group approached significance ($F = 4.01$; $df = 1/39$; $P < .10$).

Discussion

The findings of the present experiment are consistent with those of previous verbal conditioning studies (e.g., Cohen et al., 1954; Taffel, 1955) in that when a brief and general postexperimental interview was used to investigate *S*'s awareness of a correct contingency, there was evidence for conditioning without awareness. When a more extended and specific interview was employed, however, the evidence for conditioning without awareness was largely accounted for by *S*s who had been aware but whose awareness was not revealed by the brief interview. The *S*s who were inferred to be unaware on the basis of their responses to the extended interview did not, as a group, show conditioning. Thus, when *S*s were classified according to their awareness of a correct contingency the results of the present study failed to provide support for the contention that the effects of a reinforcer can be "entirely unconscious and automatic" (Dollard & Miller, 1950, p. 44). The importance of *S*'s awareness was further demonstrated by the finding that what *S* learned was a function of the nature of his awareness. The *S*s who were aware of the reinforcement of only one of the reinforced pronouns showed conditioning for that pronoun only.

When the Unaware group was separated into two groups on the basis of awareness of the *reinforcer*, however, it was found that the 13 *S*s who had been unaware of the reinforcer, in addition to being unaware of a correct contingency (Unaware-UG group), showed as much conditioning as *S*s who had been aware of a correct contingency. This finding could be interpreted as evidence for conditioning without awareness. A survey of previous research yielded little information on the performance of *S*s who were unaware of the reinforcer. Unaware *S*s in these studies are usually not interviewed carefully enough to permit an analysis of performance as a function of awareness of the reinforcer. The development of reliable interviewing procedures designed to distinguish between awareness of the reinforcer and awareness of a correct contingency would appear to be of some importance in verbal conditioning research.

An alternative to the interpretation of the performance of the Unaware-UG group as evidence for conditioning without awareness is suggested by an examination of the manner in which *S*s in this group were interviewed. Ten of the 13 *S*s who were unaware of the reinforcer were not asked any further questions beyond Question 10 since these *S*s answered this question by indicating that *E* had said nothing during the conditioning trials. It is possible that these *S*s misunderstood Question 10 and might have verbalized a correct contingency had they been interviewed further. Thus, the question of whether or not the Unaware-UG group conditioned without awareness would appear to rest, in part, upon the adequacy of the interviewing of *S*s in this group.

Whenever *S*'s response to a postexperimental interview is employed as the basis for inferring the content of his awareness during the conditioning trials, as was the case in the present study, it can be argued that an *S* who verbalizes a correct contingency might have had the contingency suggested to him by some aspect of the interview and therefore might not have been aware of it during the conditioning trials. Even though the interview in the present study was constructed so as to avoid suggesting a correct contingency and despite the fact that all of the Aware *S*s stated that they had been aware of the contingency during the conditioning trials, it is still possible that suggestion was operating. The suggestion could have been mediated by minimal verbal and nonverbal cues emitted by *E* during the interview as a result of his knowledge of *S*'s conditioning performance and of the correct contingency. This form of suggestion could be eliminated by an experiment in which two *E*s were employed. Groups of *S*s would be reinforced on different pronouns by one *E* and interviewed by another *E* who would

not know on what pronoun *S* had been reinforced or whether or not he had conditioned.

Another manner in which suggestion might have been operating is through the nature of the interview questions. The content of the questions might have stimulated *S* to look back over the task, reorganize his impressions and thereby formulate a correct contingency. In other words, *S* might have been led to verbalize a correct contingency merely as a result of being questioned about the task. This form of suggestion would be difficult to eliminate since it could always be argued that *S*'s response to any question is suggested to him by the content of the question and does not reflect something of which he was aware before the question was asked. Clearly, it is impossible to *know* what *S* was aware of during the conditioning task. However, a fruitful approach might be to attempt to elicit as much information as possible from *S* regarding his awareness during the task while at the same time taking all possible precautions to prevent suggestion from operating. The information elicited from *S* under such circumstances could well lead to a better understanding of verbal behavior.

Summary

The primary purpose of the present study was to evaluate the extent to which previous evidence for conditioning without awareness is an artifact of insensitive interviewing procedures. Seventy-nine patients were assigned to Experimental and Control groups and required to construct sentences beginning with one of six personal pronouns and containing a past tense verb. For the Experimental group, *E* reinforced sentences beginning with *I* and *we* by saying "good"; the Control group performed without receiving verbal reinforcement. A postexperimental interview was employed to obtain information regarding *S*'s awareness of a correct contingency between the reinforced response and the reinforcer. The first four questions of the interview were approximately the same as the questions used to investigate awareness in previous research; the remaining questions were designed to obtain additional information from *S* while not suggesting a correct contingency to him. Two bases for inferring awareness were employed: (*a*) *S*'s responses to the first four questions of the interview, and (*b*) *S*'s response to the full interview.

The results indicated that there was evidence for conditioning without awareness when *S*'s responses to the first four questions were used as the basis for inferring awareness. This finding was consistent with those of previous verbal conditioning studies. When the full interview was used as the basis for inferring awareness, however, the evidence for conditioning without awareness was found to have been largely accounted for by *S*s who had been aware but whose awareness was not revealed by the brief interview. There was also limited evidence for conditioning without awareness in that a group of *S*s who were unaware of the reinforcer, in addition to being unaware of a correct contingency, showed as much conditioning as *S*s who were aware of a correct contingency.

References

Buss, A., Gerjuoy, I., & Zusman, J. Verbal conditioning and extinction with verbal and non-verbal reinforcers. *J. exp. Psychol.*, 1958, **56**, 139–145.

Cohen, B. D., Kalish, H. I., Thurston, J. R., & Cohen, E. Experimental manipulation of verbal behavior. *J. exp. Psychol.*, 1954, **47**, 106–110.

Cushing, M. C. Affective components of the response class as a factor in verbal conditioning. Unpublished doctoral dissertation, Univer. Nebraska, 1957.

Dailey, J. M. Verbal conditioning without awareness. Unpublished doctoral dissertation, State Univer. Iowa, 1953.

Dollard, J., & Miller, N. E. *Personality and psychotherapy.* New York: McGraw-Hill, 1950.

Essman, W. B. "Learning without awareness" of responses to perceptual and verbal stimuli as a function of reinforcement schedule. *Percept. mot. Skills*, 1959, **9**, 15–25.

Foster, V. H., & Goddard, H. N. The Ohio Literacy Test. *Pedag. Sem.*, 1924, **31**, 340–351.

Greenspoon, J. The reinforcing effect of two spoken sounds on the frequency of two responses. *Amer. J. Psychol.*, 1955, **68**, 409–416.

Grossberg, J. M. The effect of reinforcement schedule and response class on verbal conditioning. Unpublished doctoral dissertation, Indiana Univer., 1956.

HARTMAN, C. H. Verbal behavior of schizophrenic and normal subjects as a function of types of reinforcement. Unpublished doctoral dissertation, State Univer. Iowa, 1955.

KIRMAN, W. J. The relationship of learning, with and without awareness, to personality needs. *Dissertation Abstr.*, 1958, **19**, 362.

KLEIN, S. Conditioning and extinction of operant verbal behavior in neuropsychiatric hospital patients. Unpublished doctoral dissertation, Indiana Univer., 1954.

LEVENTHAL, A. M. The effects of diagnostic category and reinforcer on learning without awareness. *J. abnorm. soc. Psychol.*, 1959, **59**, 162–166.

LINDQUIST, E. E. *Design and analysis of experiments in psychology and education.* Boston: Houghton Mifflin, 1953.

TAFFEL, C. Anxiety and the conditioning of verbal behavior. *J. abnorm. soc. Psychol.*, 1955, **51**, 496–501.

THORNDIKE, E. L., & LORGE, I. *The teacher's word book of 30,000 words.* New York: Columbia Univer. Press, 1952.

VERPLANCK, W. S. The control of the content of conversation: Reinforcement of statements of opinion. *J. abnorm. soc. Psychol.*, 1955, **51**, 668–676.

(Received December 22, 1959)

Journal of Experimental Psychology
1961, Vol. 61, No. 1, 76-81

PROACTIVE AND RETROACTIVE EFFECTS OF OVERLEARNING[1]

GEORGE MANDLER
University of Toronto

AND

CLEMENTINA K. KUHLMAN
Harvard University

A variety of studies of transfer have shown that overlearning tends to decrease retroactive or proactive interference effects of intervening or preceding activities. For example, Reid (1953) demonstrated that overlearning in a Y maze facilitated response reversal, while previously Melton and Irwin (1940) had been able to report a *decrease* in retroactive inhibition with very high degrees of interpolated learning. In reviewing some of these studies, Mandler (1954a) has suggested that overlearning results in a more efficient response to situations which evoke competing response tendencies. Subsequent studies have tended to support such an interpretation (Mandler, 1954b; Mandler & Heinemann, 1956).

In the present study we will be concerned with the effect of overlearning of a complex motor response (motor patterns) on the learning and retention of another, but similar, response. The following predictions were made: (*a*) An overlearned pattern should be more easily recovered, following an intervening task, than a pattern learned to mastery only. (*b*) An overlearned pattern will produce less of a retroactive interference effect on a prior pattern than an intervening pattern learned only to mastery. (*c*) Overlearned patterns will interfere less with subsequent new patterns than will patterns which are learned merely to mastery.

[1] This research was supported in part by a grant from the Foundations' Fund for Research in Psychiatry. It was completed while the first author was a Fellow at the Center for Advanced Study in the Behavioral Sciences.

METHOD

Task and apparatus.—The motor task chosen consisted of sequences (or patterns) of 8 switches on a switchboard of 64 spring-return lever switches arranged 2 in. apart in an 8 × 8 array. A trial consisted of the completed choice of the 8 correct switches; no time limit was used. All *Ss* were required to learn one pattern (Task 1) after which they learned a new pattern (Task 2), relearned the first pattern (Task 3), and relearned the second pattern (Task 4). Groups of *Ss* varied in the amount of training they received on each task.

At the far side of the 8 × 8 board was an array of eight lights (success signals) across the width of the board. Behind the response board and hidden from *S* was a switchboard, connected by a series of relays with the eight success signals, which permitted the programing of any sequence of eight levers. In the course of a trial the first success signal (on the far left) was lighted when *S* activated the first correct switch, the next signal was lighted when he activated the second correct switch, and so forth. Following the activation of the eighth correct switch and the lighting of the eighth (far right) signal, all lights were automatically turned off and the board was ready for the next trial. Incorrect responses did not alter the signal lights, nor did the activation of a correct response in the wrong sequence affect the signal lights. No activation of the second, third, etc., signal light was possible unless the preceding switch, or switches, had been activated. The *S*'s task was to light the eight signal lights, which was possible only if the eight corresponding switches were activated in the proper sequence, regardless of intervening errors. Eventually *S* could light the eight signal lights by activating the eight correct switches in the proper sequence.

A record of number of responses during each trial and length of each trial to the nearest .1 sec. was obtained for each *S*.

Subjects.—The *Ss* were 50 volunteer undergraduates obtained through the university employment service. They were assigned at random to 10 groups of 5 *Ss* each (five training conditions and two patterns within each condition).

Design.—Table 1 shows the five different training conditions. Each *S* was given four tasks such that the pattern he was required to learn was the same on Tasks 1 and 3 and on Tasks 2 and 4. Two patterns (A and B), shown in Fig. 1, were used. Half the *Ss* in each training group started with Pattern A on Task 1, half started with Pattern B; *Ss* who were given Pattern A on Tasks 1 and 3 were given Pattern B on Tasks 2 and 4 and vice versa. The two patterns were obtained from a table of random numbers with the restriction that no switch on Pattern A was duplicated on Pattern B.

TABLE 1

DESIGN OF THE EXPERIMENT

Group	Task 1 L Pattern 1	Task 2 L Pattern 2	Task 3 RL Pattern 1	Task 4 RL Pattern 2
I	OL	M	M	M
II	M	OL	M	M
III	M	M	M	M
IV	OL	OL	M	M
V	OL	M	OL	M

Note.—L = learn; RL = relearn; M = learning to mastery (five successive errorless trials); OL = overlearning (mastery plus 50 trials).

For the purposes of the experiment, and on the basis of preliminary studies, it was decided to establish the criterion of mastery at five successive errorless trials, and the criterion of overlearning at 50 trials, regardless of errors, beyond the criterion of mastery. Fifty additional trials were assumed to be sufficient to establish overlearning and not too many to produce unwanted fatigue. The number of overlearning trials was held constant since only a few *Ss* made errors during overlearning.

RESULTS

Three measures were obtained on the performance of the task, and the following statistical analyses were performed:

The first measure was the number of trials to a total of three errorless performances (*criterion* measure). These data are shown in Table 2. Analyses of variance were computed for the five groups on each of the four tasks.

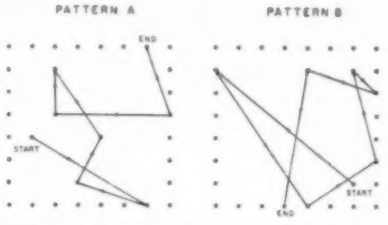

FIG. 1. Schematic presentation of the two patterns on the 64-switch board. Sequence of responses is indicated by arrows.

The second measure was the *speed* measure, which consists of the reciprocal of time in seconds per trial. The results are shown in Fig. 2. Analyses of variance were computed for the first five trials on Task 3 and 4 and for Trials 2–6 on Tasks 1 and 2. Trial 1 on Tasks 1 and 2 was eliminated because the behavior of all groups in both tasks indicated that this trial consisted of a search for and identification of the correct switches. An analysis of variance of the speed measure for Trial 1 showed $F < 1$ both between and within Tasks 1 and 2.

The third measure was the *response* measure which consists of number

TABLE 2

MEAN NUMBER OF TRIALS TO CRITERION

Group	Task 1		Task 2		Task 3		Task 4
I	13.4		11.9		3.8		6.4
						4.6	
V	14.6		12.3	12.8	5.5		6.0
IV	13.4	14.5	14.1		6.5		3.9
II	14.6		12.8		7.3		5.9
				12.2			
III	16.3		11.7		7.8		6.8
σ_m	2.84		4.06		1.50		1.45

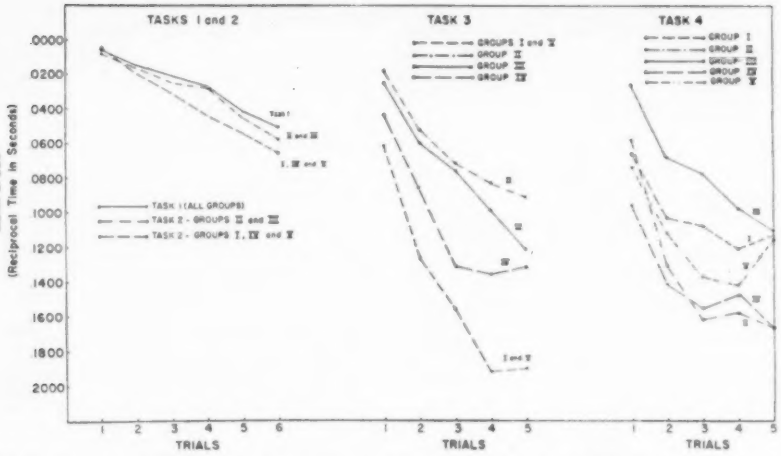

FIG. 2. Mean speed (reciprocal of time in seconds) per *S* by trials for all four tasks. Note that the ordinate is reversed in order to facilitate comparison with Fig. 3.

of responses per trial (8 responses indicate errorless performance). The results are shown in Fig. 3. Analyses of variance were performed on Trials 2–6 for Tasks 1 and 2 for the same reasons discussed for the speed measure. As seen in Fig. 3, Tasks 3 and 4 show no differences among groups beyond Trials 2 and 1, respectively. Therefore, the analyses of variance were restricted to Trials 1 and 2 on Task 3 and Trial 1 on Task 4.[2]

Tasks 1 and 2.—The criterion measure showed no significant differences either within the two tasks or between them. On the speed and response measures no significant differences were found among groups on Task 1, indicating that the groups were initially comparable. On Task 2 the analyses also fail to yield significant differences between the two aggregations which were treated identically up to this point (Groups II and III vs. Groups I, IV, and V). However, the difference between Task 1 and Task 2 (all groups combined) is significantly different on the speed measure ($P < .05$) and in the same direction ($P < .10$) on the response measure.

Task 3.—The criterion measure is significant at only the .07 level. On the speed measure we find a significant difference among groups ($P < .001$) and a significant interaction between Trials and Groups ($P < .05$). Tukey's gap test permits us to distinguish two aggregations at the .05 level: Groups I and V vs. Groups IV, III, and II. An analysis of the interaction indicates that it is a function of increasing variation among groups over trials; groups differ significantly among one another only on Trials 3, 4, and 5.

On the response measure, differences among groups are significant at the .01 level, and again the Groups by Trials interaction is significant ($P < .01$). Tukey's gap test segre-

[2] None of our analyses showed any significant main or interactive effects associated with Patterns A and B. Therefore, no reference will be made to this aspect of our design.

gates Groups I, V, and IV vs. Groups III and II. The interaction is a function of decreasing variability among Groups ($P < .05$ on Trial 1 and $> .10$ on Trial 2).

Task 4.—The criterion measure fails to show any significant differences. On the speed measure we again find a significant source of variation in Groups ($P < .01$) and a significant Groups by Trials interaction ($P < .05$). While the gap test permits no segregation for the Groups, Group III differs significantly from Groups V, IV, and II ($P < .05$ or smaller) and Group IV is significantly better than Groups I and III ($P < .05$ and $< .01$, respectively). The interaction effect is clearly contributed by Group II, which is the second slowest on Trial 1 (after Group III) and is significantly slower than Group IV ($P < .05$); but by Trials 3, 4, and 5, Group II is significantly better than Groups I, III, and V ($P < .05$ or smaller) and no different from Group IV.

On Trial 1 of the response measure the groups do not significantly differ from one another.

Discussion

The data from Tasks 1 and 2 show some facilitative effects of prior experience (on Task 1). However, mere mastery training does not produce significantly less positive transfer than does overlearning.

On Task 3 the *S*s who had overlearned the task during Task 1 (Groups I and V) perform significantly better than those who learned that task to mastery only (Groups II and III). Group IV, which had also overlearned Task 1, but overlearned the intervening task as well,

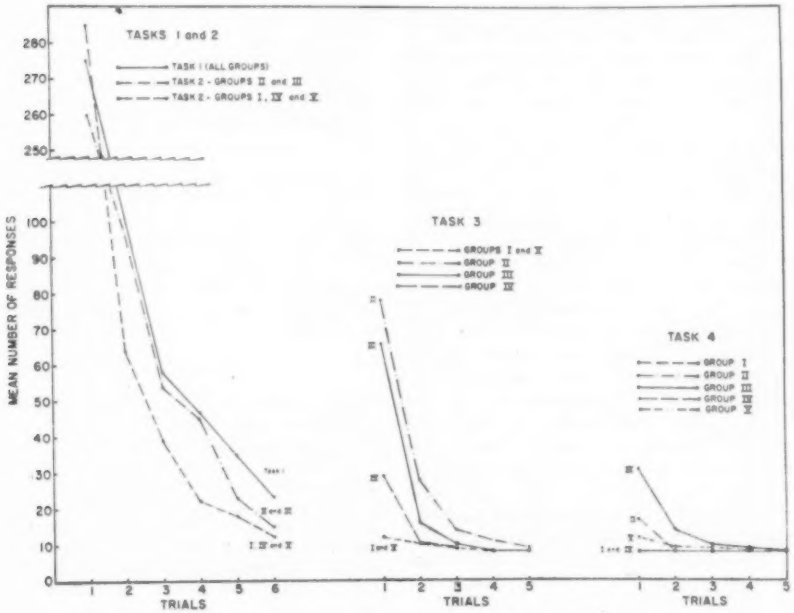

FIG. 3. Mean number of responses per *S* by trials for all four tasks.

shows behavior intermediate between these two effects; Group IV is classed with the overlearned groups on the response measure and with the mastery groups on the speed measure. Our expectation that overlearned patterns would be more easily recovered is confirmed, but the prediction that intervening overlearned patterns should interfere *less* with a prior pattern is clearly false.

On Task 4 we had predicted that the Groups should be ordered as follows: Group IV, which had overlearned both patterns; Group II, which had overlearned the pattern required in Task 4; Group V, which had overlearned the other pattern twice; Group I, which had overlearned the other pattern once; and Group III, which had been given no overlearning. This ordering was obtained for both the speed and the criterion measure (the probability of obtaining this predicted order is .008); on the response measure Groups I and II were reversed. The analyses of variance on the speed measure indicate that having had no overlearning at all (Group III) produces inferior performance to having overlearned the required pattern (Groups II and IV) or having overlearned the other pattern more than once (Group V). A comparison of Groups II and IV suggests that the advantage for Group IV in having overlearned the other pattern appears only on the first trial, thereafter Group II performs just as well.

The findings on Tasks 3 and 4 can be summarized as follows: the postulated protective and conserving effects of overlearning did not appear in the ABA paradigm, but they did appear in the ABAB paradigm. One explanation—alternative and supplementary to the overlearning hypothesis—is that the ABAB paradigm permits better discrimination between and differentiation of the two tasks than does the ABA situation where the *Ss* have had only one encounter with each task. Coupled with a hypothesis of greater task differentiation as a function of overlearning, our findings would be equally comprehensible from such a point of view.

Two other results deserve mention. It will be seen from Fig. 2 and 3 that the response measure shows a general improvement from Task 2 to Task 3 ($P < .05$), while no such improvement can be found on the speed measure. This suggests that while *Ss* make fewer errors on Task 4 generally, they behave in a relatively cautious and slow manner.

In a previous study (Mandler, 1954b), postexperimental interviews had indicated that *Ss* who were given minimal training on a task similar to the one used in the present study were unable to reproduce the correct response pattern in the absence of the response board, *Ss* with intermediate degrees of training seemed to depend on tactual cues but not the visual stimulus of the board to reproduce the response pattern, while *Ss* who had been given overlearning reported visual imagery of the response pattern independent of the visual stimulus of the board and relatively free of tactual cues. We expected that in the present study *Ss* who had had some overlearning experience would tend to describe their behavior and method primarily in terms of visual imagery, while *Ss* who had been given no overlearning would tend to use tactual imagery or cues. Chi square analysis of a comparison of the 10 *Ss* who had no overlearning (Group III) with the other 40 *Ss* showed that a lower proportion of the *Ss* in Group III made reference to visual imagery ($\chi^2 = 2.76$, $P < .10$), while a significantly higher proportion of these *Ss* made reference to tactual cues or imagery ($\chi^2 = 6.89$, $P < .01$).

Summary

Five groups of 10 *Ss* each were required to perform four tasks. On Task 1 they learned a motor pattern consisting of a sequence of 8 switches on a 64-unit switchboard. On Task 2 they learned a different pattern on the same board. On Tasks 3 and 4 they relearned the first and second patterns, respectively. The five groups differed in the amount of training they received on each task; they either learned the pattern to mastery (five successive errorless trials) or were given 50 trials beyond mastery.

Following the acquisition of one pattern, learning of a second pattern was superior to learning of the first. Overlearned patterns were more easily recovered than patterns learned to mastery only. Overlearning of an intervening pattern increased retroactive interference on Task 3; it facilitated performance on Task 4.

Interviews with *Ss* also suggested that while overlearning tended to produce more visual imagery, lower levels of learning produced more tactual imagery.

References

MANDLER, G. Response factors in human learning. *Psychol. Rev.*, 1954, **61**, 235–244. (a)

MANDLER, G. Transfer of training as a function of degree of response overlearning. *J. exp. Psychol.*, 1954, **47**, 411–417. (b)

MANDLER, G., & HEINEMANN, S. H. Effect of overlearning of a verbal response on transfer of training. *J. exp. Psychol.*, 1956, **51**, 39–46.

MELTON, A. W., & IRWIN, J. M. The influence of degree of interpolated learning on retroactive inhibition and the overt transfer of specific responses. *Amer. J. Psychol.*, 1940, **53**, 173–203.

REID, L. S. The development of noncontinuity behavior through continuity learning. *J. exp. Psychol.*, 1953, **46**, 107–112.

(Received December 28, 1959)

Journal of Experimental Psychology
1961, Vol. 61, No. 1, 82–88

EFFECTS OF INSTRUCTIONS ON THE EXTINCTION OF A CONDITIONED FINGER-WITHDRAWAL RESPONSE [1]

RICHARD H. LINDLEY AND K. E. MOYER

Carnegie Institute of Technology

Previous human conditioning studies have found that the instructions read to *Ss* affect the rate of acquisition of the conditioned eyeblink response (McAllister & McAllister, 1958; Miller, 1939), of the conditioned salivary response (Razran, 1935), and of the conditioned GSR (Cook & Harris, 1937). Other studies have shown that instructions are an important variable in transfer (Lindley, 1959), and in a conditioned discrimination (Hilgard & Humphreys, 1938).

Instructions also affect the extinction of CRs. Cook and Harris (1937) found that the conditioned GSR extinguished much faster when *Ss* were informed that the UCS would not be delivered on subsequent trials than when *Ss* were not so informed. Norris and Grant (1948) found that instructing *Ss* to try to inhibit an eyelid CR produced rapid extinction. The purpose of the present study was to investigate further the effects of instructions on the extinction of the conditioned finger-withdrawal response and to suggest an explanation of the results in terms of S-R reinforcement theory.

According to Hull (1943) drive (D) level and habit strength ($_sH_R$) are important determiners of effective reaction potential ($_sE_R$). Spence and Taylor (1951) have hypothesized that anxiety is part of the generalized D level of *Ss*. A hypothesis which accounts for many of the data of the effects of instructions on extinction is that the appropriate instructions act to lower D level (and thus $_sE_R$) by making the *Ss* less anxious.[2] If, after a CR has been established, *Ss* are informed that the UCS will no longer be delivered, anxiety (and thus D) should be reduced which would lead to a lowered $_sE_R$ and to rapid extinction. One hypothesis tested in the present study was that the conditioned finger-withdrawal response would extinguish faster when *Ss* were informed that the UCS would not be delivered during extinction than when *Ss* were not so informed. While this hypothesis has been confirmed for an autonomic response like the GSR (Cook & Harris, 1937), there seems to be no comparable study for skeletal muscle CRs.

It should also be possible to offset somewhat the effects of instructions on extinction by increasing the degree of conditioning. As the number of reinforcements is increased, $_sH_R$ is increased which means that a high degree of conditioning will produce a higher $_sH_R$ and hence a higher $_sE_R$ than a low degree of conditioning. Assuming that the appropriate in-

[1] This research was supported in part by a grant from the Humanistic and Social Studies Division of Carnegie Institute of Technology. We would like to thank Richard B. Ross for assistance in the collection and analysis of the data.

[2] It should be clear, of course, that the present hypotheses and study do not *test* Spence and Taylor's theory since the boundary conditions specified by these authors have not been met in the present study (see Spence and Ross [1959] for an extended discussion of this point). Rather, the present study attempts to *apply* S-R concepts to the complex problems involved in the effects of instructions on CRs.

structions reduce the magnitude of D by a constant amount, the high degree of conditioning will produce a higher ${}_{S}E_{R}$ at the start of extinction than the low degree of conditioning. The present study tested the hypothesis that instructions would have less effect on CRs of high ${}_{S}H_{R}$ than on CRs of low ${}_{S}H_{R}$.

METHOD

Apparatus.—The CS was a tone of 2000 cps generated by a Hewlett-Packard oscillator. The UCS was produced by an inductorium. The right hand rested palm down on the conditioning board with the middle finger extended. The finger was attached with adhesive tape to a movable ¼-in. × 2½-in. piece of plastic which contained the two electrodes. The electrodes touched the volar surface of the finger about 1 in. and 1⅛ in. proximal to the finger tip. In this position the finger was free to move up or down. The movable piece of plastic was attached to a series of contacts which recorded finger movements of ⅛ in. in discrete steps on a 10-pen Esterline-Angus operations recorder. There were four contact points for an upward movement of the finger and three for a downward movement. The operations recorder provided a graphic record on which the onset, offset, and duration of the CS and UCS were indicated as well as the latency and magnitude of finger movements.

The CS-UCS time interval was controlled by two Hunter timers. The time intervals between trials were controlled by an arrangement of a movable tape driven by a constant speed motor and a microswitch.

The S sat in a sound-deadened booth which was 4 ft. 8 in. square and 5 ft. 8 in. high. An electric exhaust fan provided both ventilation and constant noise source which further served to mask any outside noises. The booth was lighted by a 150-w. bulb which was recessed in the ceiling. The E and S were in separate rooms and communicated by means of a two-way speaker system.

Procedure.—The S was seated in a chair parallel to the conditioning board. Electrode paste was applied to the middle finger and the finger and arm were fastened in the apparatus. The arm was held in place by two flexible straps and an elbow block. The S was asked to move his finger up and down a few times and was told to hold his finger so that a pointer visible to him would be touching a designated contact point. The speaker system was explained and the rest of the instructions were presented over the speaker system.

The instructions explained the nature of the conditioning process to S. The S was cautioned not to respond voluntarily but also not to prevent his finger from moving. After the instructions were read, S repeated them in his own words and any errors or omissions were corrected by E.

The shock level was adjusted by delivering the discrete shocks of .2-sec. duration to the finger in gradually increasing intensities. Shock level was increased until the finger moved vigorously and S reported discomfort. In some cases it was necessary to increase the shock level during conditioning; however, no change in shock intensity was made after Ss reached the criterion of four CRs in five consecutive trials.

Three tests with tone alone were given. In the few cases that an S responded, the tests were continued until no responses occurred for three consecutive trials.

Classical conditioning was used. The duration of the CS was .5 sec. and it was presented .5 sec. before the UCS, which was of .2-sec. duration. The CS and UCS did not overlap. In each block of five conditioning trials, one test trial was given during which the UCS was not delivered. The test trials were presented in a prearranged haphazard sequence. The paired presentations of the CS and UCS were separated by intervals of 20–40 sec., in a prearranged random sequence.

After S reached a criterion of four CRs in five consecutive trials, he was assigned to one of the eight groups by a matching procedure. The matching variable was the number of trials to reach criterion. Half of the Ss were extinguished immediately and half were given an additional 20 conditioning trials in which one test trial was presented in a prearranged haphazard sequence in every block of five trials. After the additional 20 conditioning trials were given, Ss were extinguished. For all groups, the CS was presented alone for 25 trials during extinction. The presentations of the CS were separated by intervals of 20 to 40 sec. in a prearranged random sequence. During extinction, any responses which occurred within 1.5 sec. after the CS was presented were considered to be extinction responses. The data for CRs in extinction, i.e., responses which occurred .5 sec. or less after the CS was presented, were also analyzed.

Experimental design.—A 2 × 4 factorial design was used. There were four main

groups of *Ss* and, within each group, half the *Ss* were conditioned to the criterion of four CRs in five trials (Low-H groups) and half were given the additional 20 conditioning trials after reaching the criterion (High-H groups). The four main groups differed in *D* level. The procedure for Groups 1 and 2 was designed not to change *D* level at the start of extinction; the procedure for Groups 3 and 4 was designed to lower *D* level.

Group 1 served as a control; no instructions were read to *Ss* in this group before extinction was begun and the extinction procedure was begun without pause after the desired degree of conditioning had been reached. For Group 2 instructions which were designed to be neutral were given; these instructions recapitulated the salient portions of the original instructions. The *Ss* were told, "Remember to let your finger move automatically. Whenever your finger feels like jumping up, just let it jump up." A comparison of the number of responses elicited during extinction for Groups 1 and 2 would reveal if simply talking to *Ss* before extinction has any effect on *D* level.

The *Ss* in Group 3 were told, "*There will be no more shock delivered on any of the trials from now on.* Remember to let your finger move automatically. Whenever your finger feels like jumping up, just let it jump up." The *Ss* in Group 4 were told, "*There will be no more shock delivered on any of the trials from now on.* I want you to try to prevent your finger from moving when the tone is presented." The italicized portions of the instructions for Groups 3 and 4 were designed to lower *D* level. After the instructions were read to Groups 2, 3, and 4, *Ss* were asked to repeat them in their own words in order to ensure that they had been correctly understood. The above instructions were read to *Ss* over the speaker system.

Subjects.—The *Ss* were 139 men and women from Carnegie Institute of Technology. Seven *Ss* were eliminated because the shock did not elicit consistent UCRs or because *Ss* objected to the shock level. Eighty *Ss* reached the criterion of four CRs in five consecutive trials within 50 trials; these *Ss* were assigned to the eight groups ($N = 10$ in each group) by the matching procedure described above. In addition, 20 *Ss* did not reach criterion but gave at least one CR within 50 trials; these *Ss* were assigned alternately to either Groups 1 and 2, or Group 3, and were extinguished. Thirty-two *Ss* did not condition within the 50 trials and their data are not reported here.

Results

The acquisition and extinction data for *Ss* who reached criterion within 50 trials will be considered first. Following this the results of *Ss* who gave some CRs within 50 trials but who did not reach criterion will be presented.

Acquisition.—Analysis of variance showed that there were no significant differences among the groups for the mean number of trials to reach the criterion of four CRs in five trials. (For the Low-H groups, the mean number of trials to criterion was 19.90 for Group 1, 21.10 for Group 2, 22.40 for Group 3, 21.70 for Group 4; for the High-H groups, it was 24.30 for Group 1, 21.90 for Group 2, 20.70 for Group 3, and 21.90 for Group 4). Analysis of variance showed that there were no significant differences among the groups for the mean number of CRs elicited during the acquisition period to the criterion of four CRs in five trials. (The mean numbers of CRs for the Low-H groups were 6.90 for Group 1, 5.20 for Group 2, 5.50 for Group 3, 5.90 for Group 4; for the High-H groups they were 6.50 for Group 1, 5.40 for Group 2, 5.90 for Group 3, and 5.10 for Group 4). Since *Ss* were assigned to groups by a matching procedure, no significant differences were to be expected. Analysis of variance showed no significant differences for the mean number of CRs elicited during the 20 additional conditioning trials which were given to the high-H groups (F between groups $= 1.52$, $df = 3/36$, $P > .05$).

Extinction.—Table 1 presents the Means and *SD*s for the number of extinction responses (ERs) and the number of CRs elicited during the 25 extinction trials. The statistical analyses showed that, in general, the

TABLE 1

MEANS AND *SD*s OF NUMBER OF EXTINCTION RESPONSES AND NUMBER OF CRs DURING EXTINCTION

Group	Extinction Responses				Conditioned Responses			
	Low H		High H		Low H		High H	
	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*
1	19.10	6.56	10.40	9.28	16.90	8.03	7.30	8.94
2	12.00	8.99	13.80	9.85	9.50	9.05	12.30	8.32
3	2.40	1.69	5.50	5.20	1.60	1.50	4.50	4.46
4	1.00	1.50	4.60	6.05	.50	1.20	4.30	6.07

same levels of statistical significance were obtained with the ER and CR data. There was one exception; the ER data for Groups 1 and 2 showed a nonsignificant trend while the CR data for these groups showed the same trend which was now statistically significant. In the analyses of the extinction data presented below, only the results for the ER data will be presented except in the case of Groups 1 and 2 where both ER and CR data will be presented. Whenever the ER data are presented without the corresponding CR data, it is to be understood that the same results were obtained with the CR data.

Bartlett's test revealed that the variances of the ERs were highly heterogeneous ($\chi^2 = 43.06$, $df = 7$, $P < .001$). Therefore the difference between the instructions designed to lower D (Groups 3 and 4, both Low and High H) and the procedure designed not to change D (Groups 1 and 2, both Low and High H) was tested with a nonparametric median test. The test consisted in computing the overall median number of ERs for the 80 *S*s and then determining the number of *S*s in the two categories described in the previous sentence who were below and above the overall median. An χ^2 test showed that the results were highly significant ($\chi^2 = 26.50$, $df = 1$, $P < .001$). Thus, confirming the hypothesis, instructions designed to lower D are an important determiner of the number of ERs.

Analyses of variance were performed separately on the data of Groups 1 and 2, Low and High H, and on the data of Groups 3 and 4, Low and High H. Bartlett's test revealed that the variances of the ERs for Groups 1 and 2, Low and High H, were homogeneous ($\chi^2 = 1.54$, $df = 3$, $P > .05$) and that a square-root transformation produced homogeneity of variance of the ERs of Groups 3 and 4, Low and High H ($\chi^2 = 2.76$, $df = 3$, $P > .05$). Analysis of variance of the ERs for Groups 1 and 2, Low and High H, revealed no significant differences although the interaction between H and Groups approached significance at the .05 level ($F = 4.09$, $df = 1/18$). Analysis of variance of the CRs during extinction for these groups showed that the interaction between H and Groups was significant ($F = 5.83$, $df = 1/18$, $P < .05$). Inspection of the means for these groups (see Table 1) indicates that the significant interaction is due to a large number of CRs in extinction for Group 1, Low H, and a small number for Group 1,

High H. For the neutral instructions of Group 2, the High H produced somewhat more CRs in extinction than the Low H. The ER data show trends in the same direction. Our hypotheses predicted that a high degree of acquisition training would lead to more responses in extinction than a low degree of training since ${}_sH_R$ (and hence ${}_sE_R$) should be higher with a high degree of training than with a low degree. Two possible interpretations of these data are offered in the Discussion section.

Analysis of variance of the transformed ER data of Groups 3 and 4, Low and High H, showed that only the difference between H was significant ($F = 8.45$, $df = 1/18$, $P = .01$). Thus instructions designed to lower D have more effect on CRs of Low H than on CRs of High H as was predicted by our hypothesis.

Subjects who did not reach criterion.—There were 5 *Ss* in Group 1, 5 in Group 2, and 10 in Group 3 who gave some CRs within 50 trials but who did not reach criterion. The data of Groups 1 and 2 were combined and were compared with the data of Group 3. The mean number of ERs for the combined Groups 1 and 2 was 1.00 and for Group 3 was 2.50. A t test showed that this difference was not significant ($t = .44$, $df = 18$, $P > .05$). According to our hypothesis, Group 3 (low D) should have given fewer ERs than Groups 1 and 2 (high D); however, there was a nonsignificant trend in the opposite direction. Analysis of the 50 conditioning trials for these *Ss* suggests an explanation of this result since Group 3 made more CRs (Mean = 4.30) than Groups 1 and 2 (Mean = 2.70). Thus ${}_sH_R$ was presumably higher for Group 3 than for Groups 1 and 2 which should produce an increase in the number of ERs of Group 3 as compared to Groups 1 and 2. Therefore, for Group 3, the effects of instructions on D level may have been offset by an increase in the degree of ${}_sH_R$.

Discussion

The present study is most comparable to a study by Wickens (1939) in which *Ss* did not inhibit a well-established finger-withdrawal CR when asked to do so. Wickens' study and the present study found that the appropriate instructions did not produce one trial extinction of the conditioned finger-withdrawal response. In addition, the present study shows that the appropriate instructions did produce more rapid extinction of the CR as compared to control conditions and that with the low-D groups, the effects of instructions were a function of the degree of conditioning. These findings are in accord with the hypotheses that D and ${}_sH_R$ are joint determiners of ${}_sE_R$ and that instructions can reduce D. The fact that extinction did not occur in one trial may simply be due to the fact that the instructions used in these studies did not reduce D completely.

One of our findings did not support the general hypotheses; i.e., when either no instructions or neutral instructions were given prior to extinction, the low degree of acquisition training did not lead to significantly fewer responses in extinction than the high degree of training. In fact, a trend in the opposite direction appeared for the groups in which no instructions were given. One explanation of this finding is that *Ss* in the High-H, High-D groups adapted to the shock intensity during the 20 additional conditioning trials which led to a lowering of D for these *Ss*. Thus the lowered D may have offset the increased ${}_sH_R$ so that the net effect was that ${}_sE_R$ was approximately equal for the Low- and High-H groups at the start of extinction. Of course, in Groups 3 and 4, High H, D was lowered (by hypothesis) through instructions so that any reduction in D due to adaptation to shock level during the 20 additional conditioning trials would not be a factor in the number of ERs for these groups.

On the other hand, Lewis and Duncan (1958) found that increasing the number of acquisition trials produced fewer responses in extinction when college student *Ss* were tested in a gambling situation with various schedules of intermittent reinforcement. In the pres-

ent study, the acquisition training consisted of an 80% intermittent reinforcement schedule since there was one test (nonreinforced) trial in every block of five trials. While the study by Lewis and Duncan (1958) seems fairly far removed from a classical conditioning study, it is interesting to note that both studies found a reduction in the number of extinction responses when there was an increase in the number of acquisition trials with intermittent reinforcement schedules. A number of authors (e.g., Grant & Schipper, 1952; Lewis & Duncan, 1958) have suggested that extinction under intermittent reinforcement schedules is a function of the ease of discriminating extinction trials from conditioning trials. Increasing the number of acquisition trials under intermittent reinforcement may be a way of facilitating such a discrimination, as was suggested by Lewis and Duncan (1958). Clearly, the problem of the effects of number of acquisition trials on the extinction of CRs established under intermittent reinforcement needs systematic investigation.

Another way of interpreting the results of this study is that instructions are efficacious because they change the stimulus context so that previously learned responses are elicited which compete with or modify the CR (see Logan, 1959). The results of this study offer no evidence in favor of this interpretation since instructions which would be likely to elicit an incompatible response (Group 4, "Try to prevent your finger from moving") were no more effective in producing extinction than the instructions designed simply to lower *D* level.

Summary

The present study tested the hypothesis that instructing *Ss* before extinction that the UCS would not be delivered on subsequent trials would produce rapid extinction of a finger-withdrawal CR as compared to control (noninstructed) *Ss*. A second hypothesis was that the effectiveness of this instruction would vary with the degree of conditioning. These hypotheses were derived from a theory which assumes that performance of a CR is a joint function of habit strength and drive and that instructions can lower drive. For the 80 *Ss* who reached a criterion of four CRs in five consecutive trials during acquisition of the CR, the hypotheses were confirmed ($P < .001$ for the first hypothesis, and $P = .01$ for the second). For the 20 *Ss* who did not reach criterion but who gave evidence of conditioning, the first hypothesis was not confirmed; however, the acquisition data showed that differences in the degree of habit strength were present which most likely counteracted the effects of instructions. One finding which did not support the present reasoning was that the high degree of conditioning did not lead to more extinction responses than the low degree of conditioning for the control (noninstructed) *Ss*. It was suggested either that the drive level of the *Ss* who had a high degree of conditioning might have been reduced by adaptation to the shock level or that this finding reflected the operation of a discrimination between conditioning and extinction trials by *Ss*.

References

Cook, S. W., & Harris, R. E. The verbal conditioning of the galvanic skin reflex. *J. exp. Psychol.*, 1937, **21**, 202–210.

Grant, D. A., & Schipper, L. M. The acquisition and extinction of conditioned eyelid responses as a function of the percentage of fixed-ratio random reinforcement. *J. exp. Psychol.*, 1952, **43**, 313–320.

Hilgard, E. R., & Humphreys, L. G. The effect of supporting and antagonistic voluntary instructions on conditioned discrimination. *J. exp. Psychol.*, 1938, **22**, 291–304.

Hull, C. L. *Principles of behavior.* New York: Appleton-Century, 1943.

Lewis, D. J., & Duncan, C. P. Expectation and resistance to extinction of a lever-pulling response as a function of percentage of reinforcement and number of acquisition trials. *J. exp. Psychol.*, 1958, **55**, 121–128.

Lindley, R. H. Effects of instructions on the transfer of a conditioned response. *J. exp. Psychol.*, 1959, **57**, 6–8.

Logan, F. A. The Hull-Spence approach. In S. Koch (Ed.), *Psychology: A study of a science.* Vol. 2. New York: McGraw-Hill, 1959.

McAllister, W. R., & McAllister, D. E. Effect of knowledge of conditioning upon

eyelid conditioning. *J. exp. Psychol.*, 1958, **55**, 579–583.

MILLER, J. The effect of facilitatory and inhibitory attitudes on eyelid conditioning. Unpublished doctoral dissertation, Yale Univer., 1939. (Cited in E. R. Hilgard & D. G. Marquis. *Conditioning and learning*. New York: Appleton-Century, 1940.)

NORRIS, E. B., & GRANT, D. A. Eyelid conditioning as affected by verbally induced inhibitory set and counter reënforcement. *Amer. J. Psychol.*, 1948, **61**, 37–49.

RAZRAN, G. H. S. Conditioned responses: An experimental study and a theoretical analysis. *Arch. Psychol., NY*, 1935, **28**, No. 191.

SPENCE, K. W., & ROSS, L. E. A methodological study of the form and latency of eyelid responses in conditioning. *J. exp. Psychol.*, 1959, **58**, 376–381.

SPENCE, K. W., & TAYLOR, J. Anxiety and strength of the UCS as determiners of the amount of eyelid conditioning. *J. exp. Psychol.*, 1951, **42**, 183–188.

WICKENS, D. D. A study of voluntary and involuntary finger conditioning. *J. exp. Psychol.*, 1939, **25**, 127–140.

(Received December 31, 1959)

Journal of Experimental Psychology
1961, Vol. 61, No. 1, 89–96

INFORMATION PROCESSING BEHAVIOR: THE ROLE OF IRRELEVANT STIMULUS INFORMATION [1]

ROBERT E. MORIN, BERT FORRIN, AND WAYNE ARCHER

University of Texas

The primary purpose of the present study was to determine whether human *Ss* can make their behavior independent of stimulus variation which is uncorrelated with the choice behavior to be rewarded. Specifically, can well-practiced *Ss* disregard irrelevant cues which are presented to the same modality and which occur in spatial and temporal contiguity with relevant cues in a disjunctive reaction time (RT) task?

This problem can be placed within the context of information measures and information processing behavior. Briefly stated, the amount of information per stimulus in a stimulus series is a function of the probability distribution of the possible stimulus events. In general, the greater the number of alternative events, the greater the amount of information. Similarly, response information is determined from the observed distribution of responses. Transmitted information depends on the bivariate distribution of stimulus and response elements and reflects the association or lack of independence between stimuli and responses. More complete definitions of these terms are found in Attneave (1959).

Studies by Merkel (1885) and Hyman (1953) have indicated that RT is a linear function of stimulus information expressed in bits for the special case in which response and transmitted information are each equal to stimulus information. In a recent review of information measurement and RT, Bricker (1955) proposes that the amount of information an organism must process or transmit is the crucial determinant of RT. He contends that stimulus uncertainty, as such, is not crucial and suggests that its relation to RT in the Merkel and Hyman studies is a consequence of its confounding with transmitted information. Since the amount of stimulus information places an upper bound on transmitted information (Attneave, 1959), in order to at least partially unconfound these measures it is necessary to have experimental conditions in which stimulus information exceeds transmitted information. One way of accomplishing this is to provide more information in the stimulus display than is required for the correct response. In such circumstances it may be said that some of the stimulus information is formally irrelevant.

Gregg (1954) and Archer (1954) have studied the effects of irrelevant stimulus information, but with apparently contradictory results. Gregg found that RTs for positioning a joy stick increased when irrelevant information was added to the stimulus display after practice on a simple discrimination. It is to be noted, however, that the effect was small and that *Ss* were given relatively few trials in which to learn how to filter out irrelevant stimulus information. Archer used a task which required the positioning of a series of switches to identify a pattern on an oscilloscope. Because the response employed was rather time consuming

[1] This research was supported, in part, by a grant from the University of Texas.

CONDITION	STIMULUS SLIDE				INFORMATION	
	●	■	●●	■■	STIMULUS	MAXIMUM RESPONSE
I	1				0	0
II	1	2			1	1
III	1	2	3	4	2	2
IV	1	2	1	2	2	1
V	1	2	2	1	2	1

FIG. 1. Schematic diagram of the experimental conditions.

there was no necessity for *S*s to extract all of the relevant information "at a glance." Under these conditions identification times appeared unaffected by variations in irrelevant information. It remains possible that irrelevant information will have a degrading effect on performance under more demanding conditions.

In the studies of both Gregg and Archer 1 bit of irrelevant stimulus information meant that there was one two-valued stimulus dimension which *S* could disregard; 2 bits of irrelevant information meant that *S* did not need to attend to two of the dimensions of stimulus variation. In both studies there was a congruence of formal (informational) and perceptual irrelevance. It is not necessary, however, for each bit of stimulus information to be associated with a single stimulus dimension. The average amount of information per stimulus is a function of the statistical structuring of the sequence of alternative events and does not depend directly on the perceptual properties of the events themselves. Tasks can be constructed with irrelevant information yet requiring that *S* discriminate all the physical characteristics of the stimulus. To investigate the general utility of information measures two conditions identical with respect to amount of irrelevant information but differing greatly in terms of both the perceptual demands of the task and the kinds of associations to be learned were included in the present experiment.

METHOD

Subjects.—The *S*s were 50 right-handed males from undergraduate classes at the University of Texas. Ten *S*s were randomly assigned to each of five experimental conditions.

Apparatus.—The *S* was seated approximately 2 ft. in front of a 6-in. × 6-in. viewing screen, on which could be projected any one of four slides; one circle, two circles, one square, or two squares. The *S*'s fingers rested lightly on a panel of four microswitch keys. A black screen (40 in. × 27 in.) separated *E* from *S*.

Reaction times were measured to the nearest .01 sec. by a Standard Electric timer which started when the slide viewer came on. When *S* pressed the correct key the viewer turned off and the clock stopped.

Pretraining.—In all conditions, *S*s were first given 16 simple RT trials with the single circle as stimulus. The *S* was instructed that on each trial he would be given a vocal "ready" signal, that the slide would appear 1 to 3 sec. later, and that he was to respond as rapidly as possible. After each trial *E* recorded the RT and reported it to *S*. The intertrial interval was approximately 5 sec.

Experimental conditions.—The experimental conditions are diagramed in Fig. 1. The symbols at the top of the figure refer to the four stimulus slides. The roman numerals along the left margin designate the five experimental conditions. The numbers in the body of the figure indicate the finger of the right hand which was correct when a particular slide was presented, the index finger being given the number "1." Condition I was a simple RT task. The stimulus slide was always a single circle and *S* had to respond with his index finger. With complete redundancy in both the stimulus series and the response series, there were zero bits of information in the stimulus and zero bits in the response.

In Cond. II, two slides were used, and in the last three conditions, four were used. Within each of the last four conditions all slides used appeared with equal frequency, and the first and second order sequential probabilities were as nearly equal as they could be made. Consequently, the stimulus information associated with any one of the last four conditions is $\log_2$ of the number of stimulus slides used in that condition. Thus, as shown in Fig. 1, Cond. II had 1

bit of stimulus information and the last three conditions all had 2 bits of stimulus information.

In the first three conditions, response information was equal to stimulus information, assuming errorless performance. In each case the number of slides was equal to the number of available responses, and perfect accuracy would produce a response series with the same structure as the stimulus series. Since errors can have the effect of reducing response information, the values given in the last column of Fig. 1 are to be interpreted as upper limits on response information.

In Cond. IV and V, any one of four slides was presented, but *S* was called upon to make one of only two responses. Both of these conditions had 2 bits of stimulus information but only 1 bit of response information. One bit of stimulus information was irrelevant. The two conditions, however, made quite different perceptual demands upon *S*. In Cond. IV only the *form* dimension (circle vs. square) was relevant to the correct response. The *number* of symbols on a slide (one vs. two) was irrelevant. There was no need for *S* to attend to *number* in order to make the correct response. In Cond. V, despite the fact that 1 bit of stimulus information was formally irrelevant, *S* could not consistently respond correctly unless he attended to both *form* and *number* in combination. A single circle or two squares required one response whereas one square or two circles required a different response.

Procedure.—After pretraining, *S*s were instructed with respect to their task during the experimental trials. All *S*s, except those in Cond. I, were given 12 practice trials (with no emphasis on speed) in which to learn the associations between the slides and the correct responses.

The instructions for the test period emphasized both speed and accuracy. The test period included two sessions separated

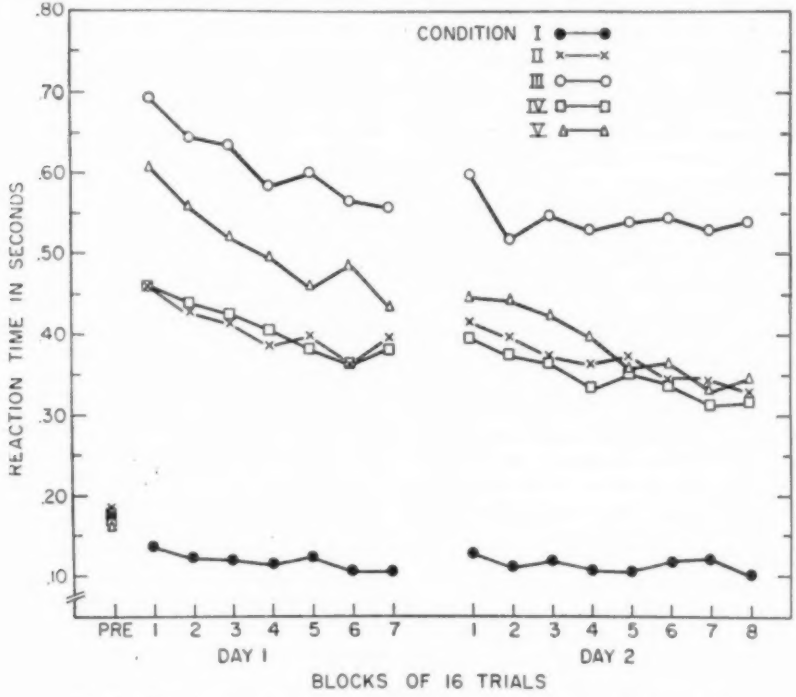

FIG. 2. Reaction time in seconds plotted as a function of successive blocks of 16 trials for the five experimental conditions.

by a 48-hr. interval. The first test session, which immediately followed the practice period, consisted of seven blocks of 16 trials; the second test session included eight such blocks. Within blocks the intertrial interval was approximately 5 sec. and there was a 1-min. rest between blocks.

As in pretraining, RTs were recorded to the nearest .01 sec. and were reported to *S*. When an error was made it was noted, but the trial was not rerun.

Results

A median RT was computed for each block of 16 trials for each *S*. Errors were regarded as extreme RTs in the determination of the medians. In Fig. 2 means of medians are plotted for each condition over blocks of trials.

On the pretraining block of trials, before the experimental treatments were applied, the average RTs for the five conditions were quite similar, ranging from a low of .164 sec. to a high of .182 sec. These differences were not significant ($F < 1.00$, $df = 4/45$).

Beginning with the first block of experimental trials and throughout the two days of practice, considerable differences in average RTs were apparent for some of the conditions. The fastest times were obtained in Cond. I, the simple RT condition with zero bits of information in the stimulus and zero bits in the response. Condition III, with 2 bits of information in the stimulus and a maximum of 2 bits in the response, yielded the slowest average RTs. The curves in Fig. 2 show the remaining three conditions to be intermediate between these extremes.

The principal comparisons among the five conditions were made in terms of terminal levels of performance. For each *S* the mean of his medians for the last four blocks of trials on Day 2 was determined. Though it would seem that these values do not necessarily represent asymptotic levels of performance, it is apparent in Fig. 2 that all conditions showed considerable leveling off by the terminal blocks of practice.

As might be anticipated from Fig. 2, an overall test of the significance of the differences in terminal performance for the five conditions was highly significant ($F = 31.10$, $df = 4/45$, $P < .01$). Of more immediate interest are certain planned comparisons among conditions. In describing these comparisons reference to response information will, at this stage, be in terms of maximum possible response information rather than obtained response information which was sometimes lower.

First, it is of interest to determine whether Hyman's basic results were replicated. In Cond. I, II, and III, as in the conditions employed by Hyman, stimulus, response, and transmitted information were all equivalent, assuming errorless performance. For this case Hyman found RT to be a linear function of stimulus information. In the present experiment the average terminal RTs for Cond. I (0 bits), II (1 bit), and III (2 bits) were .112 sec., .348 sec., and .538 sec., respectively. The data from these three conditions are consistent with Hyman's findings in that the linear regression of RT on stimulus information was highly significant ($F = 123.60$, $df = 1/45$, $P < .01$) and the curvilinear component to the regression was not significant ($F < 1.00$, $df = 1/45$).

Second, terminal performance in Cond. IV and V is of special concern. In these conditions stimulus informations (2 bits) and response information (1 bit) were not equated. As shown in Fig. 1, Cond. IV and V were identical to Cond. II with respect to response information, but like Cond. III in terms of stimulus infor-

mation. The application of Duncan's Range Test (Duncan, 1955) to the data for all five treatments indicated that mean terminal RTs for Cond. IV and V (.332 and .350 sec., respectively) did not differ significantly from each other nor from that for Cond. II (.348 sec.); they did, however, differ significantly from that for Cond. III (.538 sec.). In other words, *S*s in conditions which were alike with regard to response information, though differing with respect to stimulus information, behaved similarly. On the other hand, *S*s in conditions which differed in response information, though identical with respect to stimulus information, performed differently. The results indicate that performance was closely associated with maximum possible response information. They argue against the notion that stimulus information, as such, was a crucial determinant of RT at advanced levels of proficiency. Rather, they support the inference that *S*s were able to make their behavior independent of formally irrelevant stimulus information even, as in Cond. V, where such information was "psychologically relevant."

In contrast to Hyman's study, the amount of practice given in the present experiment (240 trials) was not sufficient to allow *S*s to achieve errorless performance. It was found, contrary to expectation, that errors were not eliminated with practice in the same orderly fashion as RTs were reduced. For Cond. II through V error rates at the end of Day 2 closely approximated those at the close of Day 1. To permit comparisons among conditions at the final stages of training, a terminal error rate was computed for each *S* by averaging his errors per block of 16 trials for the last four blocks. The mean terminal error rates for Cond. II through V were 1.35, 2.30, 2.15, and 2.88, respectively. The differences among conditions were not statistically significant ($F = 1.83$, $df = 3/36$, $P > .05$), an outcome consistent with the considerable *within-groups* variability in terminal error rates.

TABLE 1

MEASURES OF TRANSMITTED AND RESPONSE INFORMATION

Cond.	Obtained Values		Maximum Values
	Mean Transmitted	Mean Response	Transmitted and Response
I	.00	.00	.00
II	.68	.99	1.00
III	1.58	1.96	2.00
IV	.68	.99	1.00
V	.61	.97	1.00

If the errors that *S*s made were nonsystematic or random one could expect them to have different effects on transmitted and response information. Transmitted information, which is essentially a nonmetric correlation between stimulus and response events, would be lowered by such errors. On the other hand, since random errors should produce no systematic modification of the distribution of response events, response information should remain relatively unchanged. As a matter of fact, these were the results that were obtained. Measures of response and transmitted information were computed for each *S* in Cond. II through V from the data of the four terminal blocks of practice. Though stimulus and response sequences of Length 2 (diagrams) were used in these computations, the information measures reported are expressed in *bits per trial* (Attneave, 1959).

Mean transmitted and mean response information for the conditions

TABLE 2

INTERCORRELATIONS OF TRANSMITTED INFORMATION (TI), RESPONSE INFORMATION (RI), AND REACTION TIME (RT)

Cond.	TI-RI	RI-RT	TI-RT
II	.81**	.66*	.81**
III	.77**	.57	.68*
IV	.16	−.20	.75*
V	.79**	.15	−.07
I–III	.97**	.96**	.97**
I–IV	.96**	.94**	.96**
I–V	.95**	.86**	.85**

* $P < .05$.
** $P < .01$.

in question are presented in Table 1. In all conditions, mean obtained response information very closely approximated maximum possible response information. The greatest discrepancy, in Cond. III, was only .04 bits per trial. It is apparent that, when *Ss* made errors, the resulting response distributions differed minimally from those associated with errorless performance.

Errors, on the other hand, did reduce considerably the amount of information *Ss* transmitted. Table 1 shows that the mean transmitted information for the four terminal blocks of practice was only .68 bits per trial for Cond. II and IV and .61 bits for Cond. V, as contrasted with a maximum of 1.00 bit per trial for errorless performance. In Cond. III, with a maximum of 2.00 bits per trial, the mean transmitted information was 1.58 bits. The differences in amount of information transmitted in Cond. II, IV, and V were not statistically significant ($F < 1.00$, $df = 2/27$). This result is consistent with that obtained for the RT measure. With respect to both RT and transmitted information, the presence of irrelevant stimulus information, as such, had no significant influence on performance.

The presence of errors in Cond. II through V served, at least partially, to unconfound response and transmitted information, even though the correlations between these measures remained, in general, quite high. The intercorrelations of the RT, transmitted information, and obtained response information measures both within conditions and across combinations of conditions are given in Table 2.

With a single exception, the correlations of transmitted information with RT were high and statistically significant. The correlation between these measures for all 50 *Ss* in Cond. I through V, as shown in the last column and bottom row of Table 2, was .85. When the *Ss* in Cond. I through IV or I through III were considered, the correlations were .96 and .97, respectively. Even the within-groups correlations were high for Cond. II, III, and IV despite the more restricted ranges of both variables within a single condition. Only in Cond. V was the correlation between RT and transmitted information low and not significant. Examination of the scatterplot for Cond. V revealed that this outcome was associated with the performance of two highly atypical *Ss* who responded very rapidly and transmitted information with near-perfect accuracy. With the data of these two *Ss* omitted, the correlation in Cond. V was .90.

With the exception of Cond. V, the within-groups correlations of obtained response information with RT, as shown in Table 2, were slightly lower than the corresponding correlations between transmitted information and RT. Though the lower correlations are not unrelated to the more restricted within-groups variability on the response information measure the result should not be regarded as artifactual, for homogeneity with respect to response

information seems more the rule than the exception.

Discussion

Despite the fact that the amount of training given was less than that required for both asymptotic and errorless performance, *Ss* were remarkably efficient in filtering out irrelevant stimulus information. The comparison of Cond. IV and V with Cond. II indicated that performance measured in terms of RT, errors, and transmitted information did not significantly depend upon the presence or absence of irrelevant stimulus information. Even were it desirable to qualify the above statements because errors in Cond. IV and V were more frequent (though not significantly so), or because there were two superior *Ss* in Cond. V, the basic finding of efficiency in the handling of irrelevant stimulus information remains.

The general utility of describing stimulus and response events in terms of information measures depends, in part, on the breadth of conditions over which functional relations between information measures and the rate at which information is processed remain invariant. In this regard it is possible to question the generality of the finding that *Ss* can make their behavior independent of irrelevant stimulus uncertainty. The answer to this question is in turn important to the generality of Bricker's proposition that RT is determined by the amount of information an organism transmits, since his hypothesis implies that stimulus information is not an important determinant of RT.

The present study included both an indirect and a direct attempt to establish some degree of generality to its conclusions. By presenting relevant and irrelevant stimulus information to the same modality and in spatial and temporal contiguity, *S* was prevented from disregarding irrelevant stimulus information by bodily orientation or through some intersensory gating process. The fact that *Ss* performed as well as they did under these conditions permits the tentative generalization that they would be at least as efficient under conditions which offer additional possibilities for disregarding irrelevant stimulus information.

Information measures describe the statistical structuring and associations of stimulus and response events but not their perceptual properties. The usefulness of such measures in the understanding of information processing behavior might be considerably diminished if it were true that the perceptual relations among stimuli are very important determinants of the rate at which information is processed. The behavior of *Ss* in Cond. IV and V is directly relevant to the importance of perceptual factors in information processing. Though the two conditions were formally identical in terms of irrelevant stimulus uncertainty (1 bit), they were quite different with respect to the stimulus discriminations required. In Cond. V *S* had to attend jointly to both the *form* and *number* properties of the stimuli to respond correctly. Perceptually there was no irrelevant stimulus information in this condition. In Cond. IV, even if an *S* were "number blind," he would not have been handicapped because *form* was the only stimulus dimension that had to be discriminated. The results for Cond. IV and V show that the rate at which information is processed can adequately be accounted for in terms of formal information measures without reference to the perceptual demands of the task. They are consistent with the parsimonious position that RT is proportional to transmitted information.

Crossman (1953), for a condition comparable to Cond. V, found results contrary to those of the present study. He had a single *S* sort cards into two piles, red face cards and black plain cards into one pile, and black face cards and red plain cards into the other. Sorting for this condition took approximately the same amount of time as sorting the four suits into four piles and took much more time than sorting on the basis of red versus black alone or face card versus plain alone. It would

appear that sorting time was more influenced by perceptual factors than by the amount of information being transmitted. Despite the fact that Crossman's results can be subjected to the usual criticisms associated with using a single *S* as his own control, they do serve as a reminder that a number of procedural and task variables need investigation to establish general principles of information processing. The sorting task used by Crossman differed from the disjunctive RT task of the present study in that it was self-paced, used rather complex stimuli, and involved both movement and choice times. In addition, the amount of practice given *S*s must be considered.

It should be noted that, in the present study, the perceptual demands of the task did influence the rate at which information was processed during early trials. Reaction times in Cond. V exceeded those in Cond. IV on the first day of practice. In accord with Bricker (1955), Cond. V may be considered a case of suboptimal coding. Apparently, some minimum amount of time is needed to learn new or unusual codes but, once they are learned, coding may become a relatively unimportant determinant of the rate of information processing.

The data provide no striking evidence with regard to the relative importance of transmitted and response information in information processing. They are not inconsistent with the view of Hick (1952) and Bricker that RT is proportional to transmitted information, but they do not eliminate the possibility that RT is influenced by response uncertainty apart from its relation to transmitted information. It is obvious that a better technique of unconfounding response and transmitted information is needed.

Summary

The present experiment was designed to investigate the effects of irrelevant stimulus information upon information processing behavior in a disjunctive reaction time (RT) task. Ten *S*s were assigned to each of five conditions defined by the number of equiprobable stimulus events presented and by the number of correct response alternatives available. In three conditions the amount of stimulus information (0, 1, or 2 bits) and the amount of response information were equated. In two conditions stimulus uncertainty (2 bits) exceeded response uncertainty (1 bit). Though identical with respect to information measures the latter conditions differed in terms of the perceptual demands placed upon *S*.

The presence of irrelevant stimulus information did not significantly influence the rate of information processing at advanced levels of performance. This result was found to hold when (*a*) both relevant and irrelevant stimulus dimensions were presented to the same modality and in spatio-temporal contiguity, and (*b*) the character of the discrimination task required that *S* attend to all salient features of the stimulus display. On the basis of these findings it was suggested that the linear relation between stimulus information and RT observed by Hyman, and replicated in the present study, can be attributed to the confounding of stimulus, response, and transmitted information. In addition, the data provided some support for Bricker's contention that the rate of information processing is more closely associated with the amount of information transmitted than with response uncertainty.

References

Archer, E. J. Identification of visual patterns as a function of information load. *J. exp. Psychol.*, 1954, **48**, 313–317.

Attneave, F. *Applications of information theory to psychology.* New York: Holt, 1959.

Bricker, P. D. Information measurement and reaction time. In H. Quastler (Ed.), *Information theory in psychology.* Glencoe, Ill.: Free Press, 1955. Pp. 350–359.

Crossman, E. R. F. W. Entropy and choice time: The effect of frequency unbalance on choice-response. *Quart. J. exp. Psychol.*, 1953, **5**, 41–51.

Duncan, D. B. Multiple range and multiple *F* tests. *Biometrics*, 1955, **11**, 1–42.

Gregg, L. W. The effect of stimulus complexity on discriminative responses. *J. exp. Psychol.*, 1954, **48**, 289–297.

Hick, W. E. On the rate of gain of information. *Quart. J. exp. Psychol.*, 1952, **4**, 11–26.

Hyman, R. Stimulus information as a determinant of reaction time. *J. exp. Psychol.*, 1953, **45**, 188–196.

Merkel, J. Die zeitlichen Verhältnisse der Willensthatigkeit. *Phil. Stud.*, 1885, **2**, 73–127.

(Received January 2, 1960)